

90253

From: O'Hara, Eileen
Sent: Friday, March 28, 2003 6:18 PM
To: STIC-Biotech/ChemLib
Subject: sequence search for 09/803,589

SEARCH REQUEST FORM-SEQUENCE

Examiner: Eileen O'Hara (Room 10E18) Art Unit 1646

Serial Number: **09/803,589****Novel Genes Encoding Proteins Having Prognostic, Diagnostic, Preventive, Therapeutic and Other Uses
McCarthy et al.**

Date: 3/28/03 Phone: 308-3312 CM1 10A01 mailbox 10D19

Please search the following:

SEQ ID NOS: 5 and 6 (nucleic acid and encoded protein)

also, 15 consecutive nucleotides of SEQ ID NO: 5

also, 10 consecutive amino acids of SEQ ID NO: 6

in **commercial and patent** databases.

Please put results on paper.

Thank you!

Eileen B. O'Hara
CM1 10A01 mailbox 10D19
703-308-3312
U.S. Patent and Trademark Office
Art Unit 1646
eileen.o'hara@uspto.gov

POINT OF CONTACT:
PAUL SCHULWITZ
TECHNICAL INFO. SPECIALIST
CM1 6B06 TEL. (703) 305-1954

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 4/4
Date Completed: 4/14
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 12, 2003, 19:33:58 ; Search time 103 Seconds
(without alignments)
2914.918 Million cell updates/sec

Title: US-09-803-589-5

Perfect score: 979

Sequence: 1 gaattcgcacagggccagc.....gaggatccctaccctccggc 979

Scoring table: OLIGO_NUC

Gapop 60.0 , Capext 60.0

Searched: 441362 seqs, 153338381 residues

Word size : 15

Total number of hits satisfying chosen parameters: 1276

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA.*

- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	19	1.9	467	2	US-08-841-349-18
2	19	1.9	489	1	US-08-663-310-3
3	19	1.9	489	2	US-09-006-491-3
4	19	1.9	489	3	US-09-335-919-3
5	19	1.9	972	4	US-08-980-832-33
6	19	1.9	1261	4	US-08-980-832-30
7	19	1.9	1930	1	US-08-544-900-1
8	19	1.9	1930	5	PCT-US95-07874-1
9	19	1.9	2350	4	US-09-187-478-1
10	19	1.9	2350	4	US-09-292-036-1
11	19	1.9	2886	1	US-08-663-310-7
12	19	1.9	2886	3	US-09-006-491-7
13	19	1.9	2886	3	US-09-335-919-7
14	19	1.9	6854	4	US-09-194-905-7
15	19	1.9	8310	3	US-08-870-126-11
16	19	1.9	8310	4	US-09-445-247-11
17	19	1.9	33529	4	US-09-144-085-3
18	18	1.8	497	4	US-09-370-838-279
19	18	1.8	697	4	US-09-227-357-96
20	18	1.8	1230	3	US-09-188-930-25
21	18	1.8	1230	3	US-09-188-930-201
22	18	1.8	1454	2	US-08-713-000-7
23	18	1.8	1454	2	US-08-975-316-7
24	18	1.8	1454	4	US-09-211-710-7
25	18	1.8	1454	4	US-09-615-192A-7
26	18	1.8	1932	4	US-09-518-386B-2
27	18	1.8	2129	1	US-08-426-819A-34

Sequence 8, Appli
Sequence 8, Appli
Sequence 98, Appli
Sequence 2, Appli
Sequence 1, Appli
Sequence 285, App
Sequence 112, App
Sequence 148, App
Sequence 281, App
Sequence 4, Appli
Sequence 238, App
Sequence 9, Appli
Sequence 9, Appli
Sequence 9, Appli
Sequence 45, Appli
Sequence 45, Appli
Sequence 18, Appli

28 1.8 2574 2 US-08-677-734A-8
29 1.8 2574 4 US-09-097-053-8
C 30 1.8 5515 4 US-09-398-193-98
C 31 1.8 4403765 4 US-09-103-840A-2
C 32 1.8 4411529 4 US-09-103-840A-1
33 1.7 383 4 US-09-370-838-285
34 1.7 420 1 US-08-470-179-112
35 1.7 420 1 US-08-470-179-148
36 1.7 527 4 US-09-370-838-281
37 1.7 530 3 US-08-758-662-4
38 1.7 613 4 US-09-370-838-238
39 1.7 624 2 US-08-713-000-9
40 1.7 624 2 US-08-975-316-9
41 1.7 624 4 US-09-211-710-9
42 1.7 624 4 US-09-615-192A-9
43 1.7 684 2 US-08-975-316-45
44 1.7 684 4 US-09-615-192A-45
45 1.7 810 4 US-08-406-030A-18

ALIGNMENTS

RESULT 1

US-08-841-349-18/c
; Sequence 18, Application US/08841349B
; Patent No. 5955594
; GENERAL INFORMATION:
; APPLICANT: MISHRA, LOPA
; TITLE OF INVENTION: GENES CODING PROTEINS FOR EARLY LIVER DEVELOPMENT.
; FILE REFERENCE: XX/PO44700SO
; CURRENT APPLICATION NUMBER: US/08/841,349B
; CURRENT FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 467
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: For all n's in this sequence, n=(a or g or c or t)
US-08-841-349-18

Query Match 1.9%; Score 19; DB 2; Length 467;

Best Local Similarity 100.0%; Pred. No. 19;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCGGCAGGCGCAG 19
Db 466 GAATTCGGCAGGCGCAG 448
|||||

RESULT 2

US-08-663-310-3
; Sequence 3, Application US/08663310
; Patent No. 5811273
; GENERAL INFORMATION:
; APPLICANT: MISAWA, NO. 5811273ihiko
; APPLICANT: KONDO, Keiji
; APPLICANT: KAJIWARA, Susumu
; APPLICANT: YOKOYAMA, Akihiko
; TITLE OF INVENTION: DNA STRANDS USEFUL FOR THE SYNTHESIS OF
; TITLE OF INVENTION: XANTHOPHYLLS AND THE PROCESS FOR PRODUCING THE
; TITLE OF INVENTION: XANTHOPHYLLS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICANT: US/08/663,310
FILING DATE: 23-SEP-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/JP94/02220
FILING DATE: 26-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-235917
FILING DATE: 05-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 5-348737
FILING DATE: 27-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 49441/109
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 489 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..486
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 1..486
US-08-663-310-3

Query Match 1.9%; Score 19; DB 1; Length 489;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 330 GCGGCCGCGGCGCAGGAG 348
|||||
Db 462 GCGGCCGCGGCGCAGGAG 480

RESULT 3
US-09-006-491-3
Sequence 3, Application US/09006491
Patent No. 5972690
GENERAL INFORMATION:
APPLICANT: MISAWA, No. 5972690thiko
APPLICANT: KONDO, Keiji
APPLICANT: KAJIWARA, Susumu
APPLICANT: YOKOYAMA, Akihiro
TITLE OF INVENTION: DNA STRANDS USEFUL FOR THE SYNTHESIS OF
TITLE OF INVENTION: XANTHOPHYLLS AND THE PROCESS FOR PRODUCING THE
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/006,491
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/663,310
FILING DATE: 23-SEP-1996
APPLICATION NUMBER: WO PCT/JP94/02220
FILING DATE: 26-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-235917
FILING DATE: 05-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 5-348737
FILING DATE: 27-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 49441/109
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 489 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..486
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 1..486
US-09-006-491-3

Query Match 1.9%; Score 19; DB 2; Length 489;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 330 GCGGCCGCGGCGCAGGAG 348
|||||
Db 462 GCGGCCGCGGCGCAGGAG 480

RESULT 4
US-09-335-919-3
Sequence 3, Application US/09335919
Patent No. 6150130
GENERAL INFORMATION:
APPLICANT: MISAWA, No. 6150130thiko
APPLICANT: KONDO, Keiji
APPLICANT: KAJIWARA, Susumu
APPLICANT: YOKOYAMA, Akihiro
TITLE OF INVENTION: DNA STRANDS USEFUL FOR THE SYNTHESIS OF
TITLE OF INVENTION: XANTHOPHYLLS AND THE PROCESS FOR PRODUCING THE
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/335,919


```
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/663,310
; FILING DATE: 23-SEP-1996
; APPLICATION NUMBER: WO PCT/JP94/02220
; FILING DATE: 26-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6-235917
; FILING DATE: 05-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 5-348737
; FILING DATE: 27-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 49441/109
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 489 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..486
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 1..486
US-09-335-919-3
```

```
Query Match 1.9%; Score 19; DB 3; Length 489;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 330 GCGGGCGGCGGCGGCGGAG 348
|||||
DB 462 GCGGGCGGCGGCGGCGGAG 480

RESULT 5
US-08-980-832-33
; Sequence 33, Application US/08980832B
; Patent No. 6291204
; GENERAL INFORMATION:
; APPLICANT: Pasamontes, Luis
; TITLE OF INVENTION: Improved Fermentative Carotenoid Production
; FILE REFERENCE: Improved Fermentative Carotenoid
; CURRENT APPLICATION NUMBER: US/08/980,832B
; CURRENT FILING DATE: 1997-12-01
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
; LENGTH: 972
; TYPE: DNA
; ORGANISM: E-396
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Unknown
US-08-980-832-33
```

```
Query Match 1.9%; Score 19; DB 4; Length 972;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 330 GCGGGCGGCGGCGGCGGAG 348
|||||
DB 882 GCGGGCGGCGGCGGCGGAG 900
```

```
RESULT 6
US-08-980-832-30
; Sequence 30, Application US/08980832B
; Patent No. 6291204
; GENERAL INFORMATION:
; APPLICANT: Pasamontes, Luis
; TITLE OF INVENTION: Improved Fermentative Carotenoid Production
; FILE REFERENCE: Improved Fermentative Carotenoid
; CURRENT APPLICATION NUMBER: US/08/980,832B
; CURRENT FILING DATE: 1997-12-01
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 30
; LENGTH: 1261
; TYPE: DNA
; ORGANISM: Alcalligenes PC-1
US-08-980-832-30

Query Match 1.9%; Score 19; DB 4; Length 1261;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 330 GCGGGCGGCGGCGGCGGAG 348
|||||
DB 1226 GCGGGCGGCGGCGGCGGAG 1244

RESULT 7
US-08-544-900-1
; Sequence 1, Application US/08544900
; Patent No. 5736331
; GENERAL INFORMATION:
; APPLICANT: Lin, Stanley Li
; TITLE OF INVENTION: Method for Identifying Nucleic
; TITLE OF INVENTION: Acids Encoding c-fos Promoter
; TITLE OF INVENTION: Activating Proteins
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schering-Plough Corporation
; ADDRESSEE: Patent Department K-6-1 (1990)
; STREET: 2000 Galloping Hill Road
; CITY: Kenilworth
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07033-0530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7.1
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/544,900
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/272,412
; FILING DATE: 8-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Thompson, Paul A.
; REGISTRATION NUMBER: 35,385
; REFERENCE/DOCKET NUMBER: OC0439K
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 908 298 5150
; TELEFAX: 908 298 5388
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1930 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
```

```
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-544-900-1

Query Match          1.9%; Score 19; DB 1; Length 1930;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 TTTTGGTCTGCTGCTGCT 125
DB 1728 TTTTGGTCTGCTGCTGCT 1746

RESULT 8
PCT-US95-07874-1
; Sequence 1, Application PC/TUS9507874
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Method for Identifying Nucleic Acids Encoding c-fos Promoter A
; NUMBER OF SEQUENCES: 2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7.1
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07874
; FILING DATE: -June-1995
; PRIOR APPLICATION NUMBER: US 08/272,412
; FILING DATE: 8-JUL-1994
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1930 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
PCT-US95-07874-1

Query Match          1.9%; Score 19; DB 5; Length 1930;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 TTTTGGTCTGCTGCTGCT 125
DB 1728 TTTTGGTCTGCTGCTGCT 1746

RESULT 9
US-09-187-478-1
; Sequence 1, Application US/09187478
; Patent No. 6348329
; GENERAL INFORMATION:
; APPLICANT: Schmidt, Brian F.
; APPLICANT: Allen, Margaret L.
; TITLE OF INVENTION: Connective Tissue Growth (CTGF) And Methods Of Use
; FILE REFERENCE: 08766/004001
; CURRENT APPLICATION NUMBER: US/09/187,478
; CURRENT FILING DATE: 1998-11-06
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2350
; TYPE: DNA
; ORGANISM: No. 6348329mal Rate Kidney Fibroblast
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (212)..(1252)
US-09-187-478-1

Query Match          1.9%; Score 19; DB 4; Length 2350;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 TTTTGGTCTGCTGCTGCT 125
DB 1728 TTTTGGTCTGCTGCTGCT 1746

RESULT 10
US-09-292-036-1
; Sequence 1, Application US/09292036
; Patent No. 6358741
; GENERAL INFORMATION:
; APPLICANT: FIBROGEN, INC
; APPLICANT: SCHMIDT, Brian
; APPLICANT: ALLEN, Margaret
; APPLICANT: SVERDRUP, Fran
; APPLICANT: CARMICHAEL, David
; TITLE OF INVENTION: CONNECTIVE TISSUE GROWTH FACTOR (CTGF) AND METHODS OF USE
; FILE REFERENCE: FIBRO1100-1
; CURRENT APPLICATION NUMBER: US/09/292,036
; CURRENT FILING DATE: 1999-04-14
; PRIOR APPLICATION NUMBER: US 09/292,036
; PRIOR FILING DATE: 1999-04-14
; PRIOR APPLICATION NUMBER: US 09/187,478
; PRIOR FILING DATE: 1998-11-06
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2350
; TYPE: DNA
; ORGANISM: Rat
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (212)..(1252)
US-09-292-036-1

Query Match          1.9%; Score 19; DB 4; Length 2350;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 TTTTGGTCTGCTGCTGCT 125
DB 1728 TTTTGGTCTGCTGCTGCT 1746

RESULT 11
US-08-663-310-7
; Sequence 7, Application US/08663310
; Patent No. 5811273
; GENERAL INFORMATION:
; APPLICANT: MISAWA, No. 5811273hiko
; APPLICANT: KONDO, Keiji
; APPLICANT: KAJIWARA, Susumu
; APPLICANT: YOKOYAMA, Akihiko
; TITLE OF INVENTION: DNA STRANDS USEFUL FOR THE SYNTHESIS OF
; TITLE OF INVENTION: XANTHOPHYLLS AND THE PROCESS FOR PRODUCING THE
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/663,310
; FILING DATE: 23-SEP-1996
; CLASSIFICATION: 435
```

;; PRIOR APPLICATION DATA: WO PCT/JP94/02220
;; FILING DATE: 26-DEC-1994
;; PRIOR APPLICATION DATA: JP 6-235917
;; FILING DATE: 05-SEP-1994
;; PRIOR APPLICATION DATA: JP 5-348737
;; FILING DATE: 27-DEC-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: BENT, Stephen A.
;; REGISTRATION NUMBER: 29,768
;; REFERENCE/DOCKET NUMBER: 49441/109
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202)672-5300
;; TELEFAX: (202)672-5399
;; TELEX: 904136
;; INFORMATION FOR SEQ ID NO: 7:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2886 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; US-08-663-310-7

Query Match 1.9%; Score 19; DB 1; Length 2886;
Best Local Similarity 100.0%; Pred. NO. 16;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 330 GCGGGCGGCGCGCAGGAG 348
|||||
DB 1325 GCGGGCGGCGCGCAGGAG 1343

RESULT 12
US-09-006-491-7
Sequence 7, Application US/09006491
Patent No. 5972690
GENERAL INFORMATION:
APPLICANT: MISAWA, No. 5972690ihiko
APPLICANT: KONDO, Keiji
APPLICANT: KAJIWARA, Susumu
APPLICANT: YOKOYAMA, Akihiro
TITLE OF INVENTION: DNA STRANDS USEFUL FOR THE SYNTHESIS OF
XANTHOPHYLLS AND THE PROCESS FOR PRODUCING THE
TITLE OF INVENTION: XANTHOPHYLLS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/006,491
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/663,310
FILING DATE: 23-SEP-1996
APPLICATION NUMBER: WO PCT/JP94/02220
FILING DATE: 26-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-235917
FILING DATE: 05-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 5-348737

;; FILING DATE: 27-DEC-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: BENT, Stephen A.
;; REGISTRATION NUMBER: 29,768
;; REFERENCE/DOCKET NUMBER: 49441/109
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202)672-5300
;; TELEFAX: (202)672-5399
;; TELEX: 904136
;; INFORMATION FOR SEQ ID NO: 7:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2886 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; US-09-006-491-7

Query Match 1.9%; Score 19; DB 2; Length 2886;
Best Local Similarity 100.0%; Pred. NO. 16;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 330 GCGGGCGGCGCGCAGGAG 348
|||||
DB 1325 GCGGGCGGCGCGCAGGAG 1343

RESULT 13
US-09-335-919-7
Sequence 7, Application US/09335919
Patent No. 6150130
GENERAL INFORMATION:
APPLICANT: MISAWA, No. 6150130ihiko
APPLICANT: KONDO, Keiji
APPLICANT: KAJIWARA, Susumu
APPLICANT: YOKOYAMA, Akihiro
TITLE OF INVENTION: DNA STRANDS USEFUL FOR THE SYNTHESIS OF
XANTHOPHYLLS AND THE PROCESS FOR PRODUCING THE
TITLE OF INVENTION: XANTHOPHYLLS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/335,919
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/663,310
FILING DATE: 23-SEP-1996
APPLICATION NUMBER: WO PCT/JP94/02220
FILING DATE: 26-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-235917
FILING DATE: 05-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 5-348737
FILING DATE: 27-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 49441/109
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399

TELEX: 904136
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2886 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear

US-09-335-919-7

Query Match 1.9% Score 19; DB 3; Length 2886;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 330 CGGGCCGAGCGCAGGAG 348
|||||
Db 1325 CGGGCCGAGCGCAGGAG 1343

RESULT 14
US-09-194-905-7
; Sequence 7, Application US/09194905
; Patent No. 6306627
; GENERAL INFORMATION:
; APPLICANT: DECKER, Heinrich
; TITLE OF INVENTION: ISOLATION OF THE BIOSYNTHESIS GENES FOR
PSEUDO-OLIGOSACCHARIDES FROM STREPTOMYCES GLAUCESCENS
; TITLE OF INVENTION: GLA.O AND THEIR USE
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY & LARNER
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20007-5109
; COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/194,905
FILING DATE: 29-JUL-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/EP97/02826
FILING DATE: 30-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE 19622783.6
FILING DATE: 07-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Granados, Patricia D.
REGISTRATION NUMBER: 33,683
REFERENCE/DOCKET NUMBER: 026083/0193
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 6854 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)

US-09-194-905-7

Query Match 1.9% Score 19; DB 4; Length 6854;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 457 GCGAGCTCGCTCGCGCTC 475
|||||
b 2753 GCGAGCTCGCTCGCGCTC 2771

RESULT 15
US-08-870-126-11/C
; Sequence 11, Application US/08870126
; Patent No. 6048702
; GENERAL INFORMATION:
; APPLICANT: Prendergast, George C.
; APPLICANT: Sakamuro, Daitoku
; TITLE OF INVENTION: Murine and Human Box-Dependent
TITLE OF INVENTION: MYC-Interacting Protein (Binl) and Uses Therefor
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Cntr, P O Box 457
CITY: Spring House
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19477
; COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/870,126
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/435,454
FILING DATE: 05-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/652,972
FILING DATE: 24-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Rodroff, Cathy A.
REGISTRATION NUMBER: 33,980
REFERENCE/DOCKET NUMBER: WST60CUSA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 8310 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: exon
LOCATION: 680..765
OTHER INFORMATION: /note= "exon 7"
FEATURE:
NAME/KEY: exon
LOCATION: 1052..1127
OTHER INFORMATION: /note= "exon 8"
FEATURE:
NAME/KEY: exon
LOCATION: 2503..2585
OTHER INFORMATION: /note= "exon 9"
FEATURE:
NAME/KEY: exon
LOCATION: 4059..4103
OTHER INFORMATION: /note= "exon 10"
FEATURE:
NAME/KEY: exon
LOCATION: 5543..5687
OTHER INFORMATION: /note= "exon 11"
FEATURE:
NAME/KEY: exon
LOCATION: 5688..5929
OTHER INFORMATION: /note= "putative.alt.exon"
FEATURE:

; NAME/KEY: exon
; LOCATION: 7094..7221
; OTHER INFORMATION: /note= "exon 12"
US-08-870-126-11

Query Match 1.9%; Score 19; DB 3; Length 8310;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 697 GCCCGCGCGCGCTCGGC 715
|||||
Db 4280 GCCCGCGCGCGCTCGGC 4262

Search completed: April 12, 2003, 20:48:34
Job time : 139 secs



GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 12, 2003, 19:17:08 ; Search time 2722 Seconds
(without alignments)
10467.174 Million cell updates/sec

Title: US-09-803-589-5

Perfect score: 979

Sequence: 1 gaattcgacagagccagc.....gaggtacctaccctcggc 979

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 15

Total number of hits satisfying chosen parameters: 28262

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.om.*

21: em.or.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.sts.*

28: em.un.*

29: em.vi.*

30: em.htg.hum.*

31: em.htg.inv.*

32: em.htg.other.*

33: em.htg.mus.*

34: em.htg.pln.*

35: em.htg.rod.*

36: em.htg.nam.*

37: em.htg.vrt.*

38: em.sy.*

39: em.htgo.hum.*

40: em.htgo.mus.*

41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	891	91.0	993	9	BC002851	BC002851 Homo sapi
2	840	85.8	998	6	AX083428	AX083428 Sequence
3	785	80.2	969	9	AF181562	AF181562 Homo sapi
4	678	69.3	780	6	AX083418	AX083418 Sequence
5	375	38.3	113853	9	AF196971	AF196971 Homo sapi
6	286	23.1	301	6	AX336215	AX336215 Sequence
7	226	23.1	301	6	AX409054	AX409054 Sequence
8	32	3.3	945	10	AF181561	AF181561 Rattus no
9	32	3.3	975	10	BC012263	BC012263 Mus muscu
10	32	3.3	1025	10	AF293356	AF293356 Mus muscu
11	32	3.3	1027	10	AF181560	AF181560 Mus muscu
12	32	3.3	62396	2	AC098339	AC098339 Rattus no
13	32	3.3	77876	2	AC098473	AC098473 Rattus no
14	32	3.3	111948	2	AC094563	AC094563 Rattus no
15	32	3.3	151694	10	AL670169	AL670169 Mouse DNA
16	32	3.3	177613	2	AC123014	AC123014 Rattus no
17	32	3.3	177613	2	AC123014	AC123014 Rattus no
18	25	2.6	1427	5	CHRW4P	D31900 Chicken mRN
19	24	2.5	109521	2	AP005551	AP005551 Oryza sat
20	22	2.2	425	6	AX071185	AX071185 Sequence
21	22	2.2	501	10	AF155166	AF155166 Rattus no
22	22	2.2	1519	10	AF064874	AF064874 Mus muscu
23	22	2.2	3655	10	AF332063	AF332063 Mus muscu
24	22	2.2	3656	10	AF332064	AF332064 Mus muscu
25	22	2.2	3993	10	AF247453	AF247453 Rattus no
26	22	2.2	167446	2	OSJN00085	AL606682 Oryza sat
27	22	2.2	172152	2	AC115188	AC115188 Rattus no
28	22	2.2	176097	2	AC112282	AC112282 Rattus no
29	22	2.2	204050	1	AL646070	AL646070 Rattus no
30	22	2.2	214331	2	AL732312	AL732312 Mus muscu
31	22	2.2	220181	2	AC094816	AC094816 Rattus no
32	22	2.2	239758	10	AL591495	AL591495 Mouse DNA
33	22	2.2	257596	2	AC093408	AC093408 Mus muscu
34	21	2.1	533	6	AX105591	AX105591 Sequence
35	21	2.1	3737	4	AB022927	AB022927 Streptococ
36	21	2.1	7365	1	STMSVIDE0X	L37334 Streptomyce
37	21	2.1	39250	1	SVI011500	AP003864 Oryza sat
38	21	2.1	51661	8	AP003864	AP003864 Oryza sat
39	21	2.1	97310	2	AC103083	AC103083 Rattus no
40	21	2.1	139814	2	AP003964	AP003964 Rattus no
41	21	2.1	154918	9	AL138781	AL138781 Human DNA
42	21	2.1	162297	2	AC108312	AC108312 Rattus no
43	21	2.1	179931	2	AC123246	AC123246 Rattus no
44	21	2.1	202429	2	AC130160	AC130160 Rattus no
45	20	2.0	243	4	BIN421636	AJ421636 Bos indic

ALIGNMENTS

RESULT 1

BC002851

LOCUS

DEFINITION

BC002851

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

BC002851 Homo sapiens, granin-like neuroendocrine peptide precursor, clone
MGC:3466 IMAGE:3660001, mRNA, complete cds.

BC002851.1 GI:12804000

MGC.

Homo sapiens.

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 993)

Strausberg,R.

Direct Submission

993 bp mRNA linear PRI 12-JUL-2001

JOURNAL

Submitted (05-FEB-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc.mgc@nih.gov
Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S., Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R., Lim, M., Maduro, Q.L., Masiello, C., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J., Tjongson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 12 Row: n Column: 24
This clone was selected for full length sequencing because it passed the following selection criteria: GenomeScan gene prediction.

FEATURES

Location/Qualifiers
1..993

/organism="Homo sapiens"
/db_xref="LocusID:27344"
/db_xref="taxon:9606"
/clone="MGC:3466 IMAGE:3660001"
/tissue="uterus, endometrium adenocarcinoma"
/clone_lib="NH_MGC_44"
/lab_host="DH10B-R"
/note="vector: pOTB7"
30..812
/codon_start=1
/product="granin-like neuroendocrine peptide precursor"
/protein_id="AAH02851.1"
/db_xref="GI:12804001"
/translation="MAGSPLLMWPRAGVGLVLLGLFRPPALCARVKEPRGLS
AASPLAETGAPRRFRSPVGEAAGVLAALAHLEAERQERARAEAEQDQ
ARVLAQLRVGAPRNSDPALGLDDPDPAQAOLARLARLDPAALQVLVAPVP
AAALRPVVDGDPAGDAEAGDETPOVDPELLRVLIGRILAGSADSEGVAPRL
RRADHDVGSLELPEGLGALLRKRLETPAPQVPARLLPP"
BASE COUNT 136 a 381 c 343 g 133 t

ORIGIN

Query Match 91.0%; Score 891; DB 9; Length 993;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 941; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 38 GCGCGGCTCGTGGGACAGTGGCGGGTCCCGCTCTCTGGGGGCGCGGGCGGGG 97
Db 10 GCGCGGCTCGTGGGACAGTGGCGGGTCCCGCTCTCTGGGGGCGCGGGCGGGG 69
Qy 98 GCGTGGCGCTTTGGTGTCTGTCTGTCTGGCGCTTTTCGGCGCGCGCGCTCTCG 157
Db 70 GCGTGGCGCTTTGGTGTCTGTCTGTCTGGCGCTTTTCGGCGCGCGCGCTCTCG 129
Qy 158 GCGCGCGGTAAGAGAGCGCGCGCTTAAGCGAGCGCTCTCCGCGCTTGGCTGAGACTG 217
Db 130 GCGCGCGGTAAGAGAGCGCGCGCTTAAGCGAGCGCTCTCCGCGCTTGGCTGAGACTG 189
Qy 218 GCGTCTCTCGCGCTTTCGGCGGTCAGTGTCCCGAGGTGAGCGCGCGGGCGGTGACG 277
Db 190 GCGTCTCTCGCGCTTTCGGCGGTCAGTGTCCCGAGGTGAGCGCGCGGGCGGTGACG 249
Qy 278 AGCTGGCGGGGCGTGGCGCATCTGCTGGAGCGGACAGTCCAGGAGCGGGCGGGCGG 337

Db 250 AGCTGGCGGGGCGTGGCGCATCTGCTGGAGGCCAAAGTCCAGAGCGGGGCGGGCGG 309
Qy 338 AGGCCAGGAGGCTGAGGATCAGGAGCGCGTCTGGCGAGTGTGCGCGTCTGGG 397
Db 310 AGGCCAGGAGGCTGAGGATCAGGAGCGCGTCTGGCGAGTGTGCGCGTCTGGG 369
Qy 398 GCGCGCGCGCAACTCTGATCGGCTTGGGCTTGGAGACAGACCCCGCGCTTCGAG 457
Db 370 GCGCGCGCGCAACTCTGATCGGCTTGGGCTTGGAGACAGACCCCGCGCTTCGAG 429
Qy 458 GCGAGCTGCTGCGCTCTGCTCCGCGCGCGCTTGGAGCTCCGCGCTTAGAGCCGAGC 517
Db 430 GCGAGCTGCTGCGCTCTGCTCCGCGCGCGCTTGGAGCTCCGCGCTTAGAGCCGAGC 489
Qy 518 TTGTCCTCCGCGCGCTTCCGCGCGCGCTTCCGAGCCCGCGCGCTTACGAGCAGG 577
Db 490 TTGTCCTCCGCGCGCTTCCGCGCGCGCTTCCGAGCCCGCGCGCTTACGAGCAGG 549
Qy 578 GCGCGCGCGCGCGCTGCTGAGGAGCGGAGCAGACACCCCGCGCTTGGAGCCGAGC 637
Db 550 GCGCGCGCGCGCGCTGCTGAGGAGCGGAGCAGACACCCCGCGCTTGGAGCCGAGC 609
Qy 638 TGTTCAGTACTTGTGAGGAGGATCTTGGGGAAGCGGAGCTCCGAGGGGGTGGCAG 697
Db 610 TGTTCAGTACTTGTGAGGAGGATCTTGGGGAAGCGGAGCTCCGAGGGGGTGGCAG 669
Qy 698 GCGCGCGCGCTTCCGCGCGCGCTTCCGAGCGGAGTGTGGCTCTGAGTGTCCCGCTGAGG 757
Db 670 GCGCGCGCGCTTCCGCGCGCGCTTCCGAGCGGAGTGTGGCTCTGAGTGTCCCGCTGAGG 729
Qy 758 GCGTCTGGGGGCGCTGCTGTGTGAAGCGCTAGAGACCCCGCGCGCTTGGAGTGCCTG 817
Db 730 GCGTCTGGGGGCGCTGCTGTGTGAAGCGCTAGAGACCCCGCGCGCTTGGAGTGCCTG 789
Qy 818 CAGCGCGCTTGTGCGACCTTGTGAGACCTGCGCGGATCCGCTGCGAGCCCTGGGAGCCAGAG 877
Db 790 CAGCGCGCTTGTGCGACCTTGTGAGACCTGCGCGGATCCGCTGCGAGCCCTGGGAGCCAGAG 849
Qy 878 TGCCCGCGCATCCGCGCGCGCTTCCGCGCGCGCTTCCGCGCGAGCTCCAGAGCACTTACCC 937
Db 850 TGCCCGCGCATCCGCGCGCGCTTCCGCGCGCGCTTCCGCGCGAGCTCCAGAGCACTTACCC 909
Qy 938 GCGCGAGCCAGCGCTTCTCACCGGAGGATCCCTTACCCCTTGGC 979
Db 910 GCGCGAGCCAGCGCTTCTCACCGGAGGATCCCTTACCCCTTGGC 951
RESULT 2
AX083428 LOCUS 998 bp DNA linear PAT 28-FEB-2001
DEFINITION Sequence 120 from Patent WO0112660.
ACCESSION AX083428
VERSION AX083428.1 GI:13185274
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 998)
AUTHORS Kato, S. and Kimura, T.
TITLE Human proteins having hydrophobic domains and dnas encoding these proteins
JOURNAL Patent: WO 0112660-A 120 22-FEB-2001;
SAGAMI CHEMICAL RESEARCH CENTER (JP) ; Protegene Inc. (JP)
FEATURES
source Location/Qualifiers
1..998
/organism="Homo sapiens"
/db_xref="taxon:9606"
50..832
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC33311.1"


```
/db_xref="GI:13185275"
/translation="MAGSPLLMGPRAGGVGLLVLLGLFRPPPCALCARPVKEPRGLS
AASPPLAETGAPRRFRSVPGEAGAVOELARALAHLEAEQERARAEAEADQ
ARVLAQLLRVWGAPRNSDPALGDDDDPDPAQAOLARALLARLDPAALAAQVLPAPV
AAALRPVPVYDDGAPGADAEAGDETDPDPELLRYLLGRILAGSADSEGVAAAPRL
RRAADHDVGSSELPPEGLVGLALLRVKRLTTPAPQVPARLLPP"

BASE COUNT      120 a   394 c   347 g   137 t
ORIGIN

Query Match      85.8%; Score 840; DB 6; Length 998;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 940; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 38 GCCCGGCTCGCTGGGAGCATGGCGGGTCGCCGCTGCTCTGGGGGCCCGGGCCGGG 97
Db 30 GCCCGGCTCGCTGGGAGCATGGCGGGTCGCCGCTGCTCTGGGGGCCCGGGCCGGG 89
QY 98 GCGTCGGCCCTTTGGTGTCTGCTGCTGGCCCTGTTTGGCGCCGCCCGCGCTCTGG 157
Db 90 GCGTCGGCCCTTTGGTGTCTGCTGCTGGCCCTGTTTGGCGCCGCCCGCGCTCTGG 149
QY 158 CGCGCGCGGTAAAGAGAGCCCGCGGCTAAGCGCAGCGTCTCGCGCCCTTGGCTGAGACTG 217
Db 150 CGCGCGCGGTAAAGAGAGCCCGCGGCTAAGCGCAGCGTCTCGCGCCCTTGGCTGAGACTG 209
QY 218 GCGCTCTCTGCCGCTTCCGGCGGTCACTGATCCCGAGGTGAGCGGGGGCGGCTGCAGG 277
Db 210 GCGCTCTCTGCCGCTTCCGGCGGTCACTGATCCCGAGGTGAGCGGGGGCGGCTGCAGG 269
QY 278 AGCTGGCGGGCGCTGGCGCATCTGCTGAGAGCCCAAGCTCAGGAGCGGGCGGGCGCG 337
Db 270 AGCTGGCGGGCGCTGGCGCATCTGCTGAGAGCCCAAGCTCAGGAGCGGGCGGGCGCG 329
QY 338 AGCGCAGGAGGCTGAGGATCAGCAGCGCGCGCTGCTGGCGCAGCTGCTGCGGCTCTGGG 397
Db 330 AGCGCAGGAGGCTGAGGATCAGCAGCGCGCGCTGCTGGCGCAGCTGCTGCGGCTCTGGG 389
QY 398 GCGCGCCCGCCCACTCTGATCCGGCTTGGGCTTGACAGCAGCACCOCGAGCGCCCTGCAG 457
Db 390 GCGCGCCCGCCCACTCTGATCCGGCTTGGGCTTGACAGCAGCACCOCGAGCGCCCTGCAG 449
QY 458 CGCAGCTCGCTCGCTGCTCGCGCGCGCTGCTGACCCCGCGCGCTGCTGAGCAGCAGC 517
Db 450 CGCAGCTCGCTCGCTGCTCGCGCGCGCTGCTGACCCCGCGCGCTGCTGAGCAGCAGC 509
QY 518 TTGTCCCGCGCGCTGCCCGCGCGCGCTCGCAGACCCCGCGCGCGCTGCTGAGCAGCAGC 577
Db 510 TTGTCCCGCGCGCTGCCCGCGCGCGCTCGCAGACCCCGCGCGCGCTGCTGAGCAGCAGC 569
QY 578 GCGCGGGCGCGGATGCTGAGGAGCAGCGGCGCAGCAGCACCOCGAGCTGGAGCCCGCAGC 637
Db 570 GCGCGGGCGCGGATGCTGAGGAGCAGCGGCGCAGCAGCACCOCGAGCTGGAGCCCGCAGC 629
QY 638 TGTGTAGGTACTTGTGGGAGCGATCTTGTGGGAGAGCGCGGACTCCGAGGGGGTGGCAG 697
Db 630 TGTGTAGGTACTTGTGGGAGCGATCTTGTGGGAGAGCGCGGACTCCGAGGGGGTGGCAG 689
QY 698 CCGCGCGCGCGCTCCGCGCTGCGCGGACACCATGTGGGCTCTGAGCTGCCCGCTGAGG 757
Db 690 CCGCGCGCGCGCTCCGCGCTGCGCGGACACCATGTGGGCTCTGAGCTGCCCGCTGAGG 749
QY 758 GCGTGTGGGGGCGCTGCTGCTGTGAACCGCTAGAGACCCCGCGCGCGCCAGGTGCCTG 817
Db 750 GCGTGTGGGGGCGCTGCTGCTGTGAACCGCTAGAGACCCCGCGCGCGCCAGGTGCCTG 809
QY 818 CAGCGCGCTCTTGGCACCTGAGCACTGCCCGGATCCCGTGCACCTTGGGACCCAGAG 877
Db 810 CAGCGCGCTCTTGGCACCTTGGCACTGCCCGGATCCCGTGCACCTTGGGACCCAGAG 869
QY 878 TGCCCCCGGCTCCCGCAGCAGTGTCTCCCGCGCAGCAGCTCCAGAGCAACTTACCC 937
Db 870 TGCCCCCGGCTCCCGCAGCAGTGTCTCCCGCGCAGCAGCTCCAGAGCAACTTACCC 929
```

```
QY 938 CGGCCAGGCGCCTCTCACCCGAGGATCCTACCCCTCGCTACCCCTGGC 979
Db 930 CGGCCAGGCGCCTCTCACCCGAGGATCCTACCCCTCGCTACCCCTGGC 971

RESULT 3
AF181562      969 bp      mRNA      linear      PRI 27-JAN-2000
LOCUS      Homo sapiens proSAAS mRNA, complete cds.
DEFINITION      AF181562
ACCESSION      AF181562
VERSION      AF181562.1 GI:6653212
KEYWORDS
SOURCE      Homo sapiens.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 969)
AUTHORS      Fricker,L., McKinzie,A.A., Sun,J., Curran,E., Qian,Y., Yan,L.,
Patterson,S.D., Courchesne,P.L., Richards,B., Levin,N., Mzhavia,N.,
Devi,L.A. and Douglass,J.
TITLE      Identification and characterization of proSAAS, a granin-like
neuroendocrine peptide precursor that inhibits prohormone
processing
JOURNAL      J. Neurosci. 20 (2), 639-648 (2000)
MEDLINE      20098938
PUBMED      10632593
REFERENCE      2 (bases 1 to 969)
AUTHORS      Fricker,L., McKinzie,A.A., Sun,J., Curran,E., Qian,Y., Yan,L.,
Patterson,S.D., Courchesne,P.L., Richards,B., Levin,N., Mzhavia,N.,
Devi,L.A. and Douglass,J.
TITLE      Direct Submission
JOURNAL      Submitted (30-AUG-1999) Department of Molecular Pharmacology,
Albert Einstein College of Medicine, 1300 Morris Park Avenue,
Bronx, NY 10461, USA
FEATURES
source      Location/Qualifiers
1..969
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="FLHSAAS"
1..26
5'UTR
CDS
27..809
/function="inhibits prohormone processing"
/notes="granin-like neuroendocrine peptide precursor"
/codon_start=1
/product="proSAAS"
/protein_id="AAF22643.1"
/db_xref="GI:6653213"
/translation="MAGSPLLMGPRAGGVGLLVLLGLFRPPPCALCARPVKEPRGLS
AASPPLAETGAPRRFRSVPGEAGAVOELARALAHLEAEQERARAEAEADQ
ARVLAQLLRVWGAPRNSDPALGDDDDPDPAQAOLARALLARLDPAALAAQVLPAPV
AAALRPVPVYDDGAPGADAEAGDETDPDPELLRYLLGRILAGSADSEGVAAAPRL
RRAADHDVGSSELPPEGLVGLALLRVKRLTTPAPQVPARLLPP"
sig_peptide      27..125
3'UTR      810..969
BASE COUNT      117 a   379 c   338 g   135 t
ORIGIN

Query Match      80.2%; Score 785; DB 9; Length 969;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 935; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 42 GCGTCGCTGGGAGCATGGCGGGTCCGCGCTGCTCTGGGGGCCCGGGCCGGGCGGT 101
Db 11 GCGTCGCTGGGAGCATGGCGGGTCCGCGCTGCTCTGGGGGCCCGGGCCGGGCGGT 70
QY 102 CCGCCTTTTGGTGTGCTGCTGCTCGCCCTGTTTCGGCGCGCCCGCGCTCTGCGCGCG 161
Db 71 CCGCCTTTTGGTGTGCTGCTGCTCGCCCTGTTTCGGCGCGCCCGCGCTCTGCGCGCG 130
QY 162 GCGCGTAAAGAGAGCCCGCGGCTTAGCGCAGGCTCTCCGCCCTTGGCTGAGACTGGCGC 221
Db 131 GCGCGTAAAGAGAGCCCGCGGCTTAGCGCAGGCTCTCCGCCCTTGGCTGAGACTGGCGC 190
```

Qy	222	TCCTCCGCGCTTCCGGCGGTCACTGCTCCCGCCAGGCTGAGCGGGGGGGGGGGTGCAGGAGCT	281
Db	191	TCCTCCGCGCTTCCGGCGGTCACTGCTCCCGCCAGGCTGAGCGGGGGGGGGTGCAGGAGCT	250
Qy	282	GGCGGGGGCGTGGCGCATCTGCTGGAGCGCAAGCTCAGAGCGGGCGCGGGCGGAGGC	341
Db	251	GGCGGGGGCGTGGCGCATCTGCTGGAGCGCAAGCTCAGAGCGGGCGCGGGCGGAGGC	310
Qy	342	CAGAGAGCTGAGGATCAGCAGCGCGCTCTTGGCGAGCTGCTGCGCTCTGGGGCGC	401
Db	311	CAGAGAGCTGAGGATCAGCAGCGCGCTCTTGGCGAGCTGCTGCGCTCTGGGGCGC	370
Qy	402	CCCCCGCAACTCTGATCCGGCTCTGGGGCTTGGACAGACACCCGAGCGCTGCAGCGCA	461
Db	371	CCCCCGCAACTCTGATCCGGCTCTGGGGCTTGGACAGACACCCGAGCGCTGCAGCGCA	430
Qy	462	GCTGCTCGCGCTGCTGCTCCGCGCGCTTGAACCTCCGCGCTTGGAGCGCGCTGCTGT	521
Db	431	GCTGCTCGCGCTGCTGCTCCGCGCGCTTGAACCTCCGCGCTTGGAGCGCGCTGCTGT	490
Qy	522	CCCCCGCGCTTCCCGCGCGCGCTCCGACCCCGCGCGCTTACGACGAGCGGCC	581
Db	491	CCCCCGCGCTTCCCGCGCGCGCTCCGACCCCGCGCGCTTACGACGAGCGGCC	550
Qy	582	CGCGGGCGCGATGCTGAGGAGCGAGCGAGACACCCGAGCTGGACCCGAGCTGTT	641
Db	551	CGCGGGCGCGATGCTGAGGAGCGAGCGAGACACCCGAGCTGGACCCGAGCTGTT	610
Qy	642	GAGTACTTGTGGACGATTTCTGCGGGAAGCGGACCTCCGAGGGGGTGGAGCGCC	701
Db	611	GAGTACTTGTGGACGATTTCTGCGGGAAGCGGACCTCCGAGGGGGTGGAGCGCC	670
Qy	702	CGCGCGCTTCCGCGCTGCGCGACACGATGTGGGCTCTGAGCTGCCCGCTGAGGGGT	761
Db	671	CGCGCGCTTCCGCGCTGCGCGACACGATGTGGGCTCTGAGCTGCCCGCTGAGGGGT	730
Qy	762	GCTGGGGCGCTGCTGGTGTGAACGCTAGACACCCCGCGCGCTGAGCTGCACG	821
Db	731	GCTGGGGCGCTGCTGGTGTGAACGCTAGACACCCCGCGCGCTGAGCTGCACG	790
Qy	822	CCGCTCTTGGCACCTGAGCACTGCCCGGATCCGTCACCTGGGACCCAGAGTCC	881
Db	791	CCGCTCTTGGCACCTGAGCACTGCCCGGATCCGTCACCTGGGACCCAGAGTCC	850
Qy	882	CCGCGCATCCCGCACAGGACTGCTCCCGCCAGCAGCTCCAGCAACTTACCCCGGC	941
Db	851	CCGCGCATCCCGCACAGGACTTCTCCCGCCAGCAGCTCCAGCAACTTACCCCGGC	910
Qy	942	CAGCCAGCCCTCTCACCCGAGGATCCCTTACCCCTGGC	979
Db	911	CAGCCAGCCCTCTCACCCGAGGATCCCTTACCCCTGGC	948
RESULT 4			
AX083418			
LOCUS			
DEFINITION			
AX083418			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
human.			
ORGANISM			
Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
REFERENCE			
1 (bases 1 to 780)			
AUTHORS			
Kato, S. and Kimura, T.			
TITLE			
Human proteins having hydrophobic domains and dnas encoding these			
proteins			
JOURNAL			
Patent: WO 0112660-A 110 22-FEB-2001;			
SAGAMI CHEMICAL RESEARCH CENTER (JP); Protegene Inc. (JP)			
FEATURES			
Location/Qualifiers			
1..780			
/organism="Homo sapiens"			

BASE COUNT			
ORIGIN			
79 a 297 c 294 g 110 t			
/db_xref="taxon:9606"			
Query Match			
69.3%; Score 678; DB 6; Length 780;			
Best Local Similarity 99.7%; Pred. No. 0;			
Matches 778; Conservative 0; Mismatches 2; Indels 0; Gaps 0;			
Qy	58	ATGCGGGGCTGCGCGCTCTCTGCGGGCGCGGGCGCGGGCGTGGCGCTTTTGTGCTG	117
Db	1	ATGCGGGGCTGCGCGCTCTCTGCGGGCGCGGGCGCGGGCGTGGCGCTTTTGTGCTG	60
Qy	118	CTGTGCTGCGCGCTGTTTCGGCGCGCGCGCGCTCTGCGGGCGCGGGTAAAGAGGCC	177
Db	61	CTGTGCTGCGCGCTGTTTCGGCGCGCGCGCGCTCTGCGGGCGCGGGTAAAGAGGCC	120
Qy	178	CGCGGCTTAAAGCGAGCGCTCTGCGCGCTGAGACTGCGGCTTCTGCGGCTTCGG	237
Db	121	CGCGGCTTAAAGCGAGCGCTCTGCGCGCTGAGACTGCGGCTTCTGCGGCTTCGG	180
Qy	238	CGGTGCTGCGCGCTGAGCGCGGGCGGTGAGAGCTGCGCGGGCGCTGGCG	297
Db	181	CGGTGCTGCGCGCTGAGCGCGGGCGGTGAGAGCTGCGCGGGCGCTGGCG	240
Qy	298	CATGCTGAGGCGCGACGTCAGAGCGCGGGCGCGGGCGAGGCGAGAGCTGAGAT	357
Db	241	CATGCTGAGGCGCGACGTCAGAGCGGGCGCGGGCGAGGCGAGAGCTGAGAT	300
Qy	358	CAGCAGCGCGCGCTTGGCGCGAGCTGCTGCGCGCTGCGGGCGCGCGGCTCTGAT	417
Db	301	CAGCAGCGCGCGCTTGGCGCGAGCTGCTGCGCGCTGCGGGCGCGCGGCTCTGAT	360
Qy	418	CCGCTCTGCGGCTTGGAGCGACGCCGCGCTGAGCGAGCTGCTGCGCGCTCTG	477
Db	361	CCGCTCTGCGGCTTGGAGCGACGCCGCGCTGAGCGAGCTGCTGCGCGCTCTG	420
Qy	478	CTCGCGCGCGCTTGGCGCGCTTACAGCGCGCTTACAGCGCGCGCGCGCGCTG	537
Db	421	CTCGCGCGCGCTTGGCGCGCTTACAGCGCGCTTACAGCGCGCGCGCGCGCTG	480
Qy	538	GCGCGCGCGCTTGGCGCGCGCGCGCTTACAGCGCGCGCGCGCGCGCGCTG	597
Db	481	GCGCGCGCGCTTGGCGCGCGCGCGCTTACAGCGCGCGCGCGCGCGCGCTG	540
Qy	598	GAGGAGCGCGCGCGCGCGCGCGCGCGCTTACAGCGCGCGCGCGCGCGCTG	657
Db	541	GAGGAGCGCGCGCGCGCGCGCGCGCGCTTACAGCGCGCGCGCGCGCGCTG	600
Qy	658	CGGATTTCTGCGGGAAGCGGAGCTCCGAGGGGTGGCAGCGCGCGCGCGCTG	717
Db	601	CGGATTTCTGCGGGAAGCGGAGCTCCGAGGGGTGGCAGCGCGCGCGCGCTG	660
Qy	718	GCGCGCGCGCGCGCGCGCGCGCGCTTACAGCGCGCGCGCGCGCGCGCTG	777
Db	661	GCGCGCGCGCGCGCGCGCGCGCGCTTACAGCGCGCGCGCGCGCGCGCTG	720
Qy	778	CGTGTGAACCGCTTACAGCGCGCGCGCGCGCGCGCTTACAGCGCGCGCTG	837
Db	721	CGTGTGAACCGCTTACAGCGCGCGCGCGCGCGCGCTTACAGCGCGCGCTG	780
RESULT 5			
AF196971/c			
LOCUS			
DEFINITION			
113853 bp DNA linear PRI 09-NOV-1999			
Homo sapiens GATA-binding protein 1 and histone deacetylase-like			
protein genes, complete cds; CRAS pseudogene, complete sequence;			
and protein translocase gene, partial cds.			
ACCESSION			
AF196971			
VERSION			
AF196971.1 GI:6289080			
KEYWORDS			
HTG.			
SOURCE			
Homo sapiens.			
ORGANISM			
Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			

```

REFERENCE      Mammalia: Eutheria: Primates; Catarrhini; Hominidae; Homo.
AUTHORS        1 (bases 1 to 113853)
               Blechschmidt,K., Nyakatura,G., Strom,T.M., Drescher,B., Menzel,U.,
               Meindl,A. and Rosenthal,A.
TITLE          Direct Submission
JOURNAL        Submitted (14-OCT-1999) Genome Analysis, Institute of Molecular
               Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
COMMENT        1-36552: cosmid J1045; 11106-52106: cosmid E2237; 27017-67034:
               cosmid L1850; 58536-92644: cosmid D1425; 80573-113853: cosmid
               E1239.
FEATURES       Location/Qualifiers
               1. .113853
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /chromosomes="X"
               /map="Xp11.23"
               /clone="cosmids J1045; E2237; L1850; D1425; E1239"
               1. .252
               /evidence=not_experimental
               /rpt_family="AluSx"
               254. .396
               /evidence=not_experimental
               /rpt_family="L1M4"
               397. .704
               /evidence=not_experimental
               /rpt_family="AluSg"
               705. .717
               /evidence=not_experimental
               /rpt_family="L1M4"
               817. .1083
               /evidence=not_experimental
               /rpt_family="L1MB7"
               complement(1084. .1371)
               /evidence=not_experimental
               /rpt_family="AluJo"
               1372. .1790
               /evidence=not_experimental
               /rpt_family="L1MB7"
               1802. .1924
               /evidence=not_experimental
               /rpt_family="L1MD2"
               complement(2269. .2440)
               /evidence=not_experimental
               /rpt_family="L2"
               complement(2950. .3059)
               /evidence=not_experimental
               /rpt_family="MIR"
               3105. .3235
               /evidence=not_experimental
               /rpt_family="MIR"
               3268. .3472
               /evidence=not_experimental
               /rpt_family="AluJo"
               3504. .3608
               /note="homology = 80.00%, score = 28, counts = 3"
               /evidence=not_experimental
               /rpt_type=tandem
               /rpt_unit=ttagcattagaatagtttactattataaattagta
               3613. .3955
               /evidence=not_experimental
               /rpt_family="L1M4S"
               3964. .4059
               /note="homology = 77.10%, score = 20, counts = 3"
               /evidence=not_experimental
               /rpt_type=tandem
               /rpt_unit=aataattagttactactaataagtatcagtat
               complement(4121. .4369)
               /evidence=not_experimental
               /rpt_family="MIR"
               4393. .4463
               /evidence=not_experimental
               /rpt_family="L2"
               complement(4502. .4544)
               /evidence=not_experimental
               /rpt_family="MIR"

               /evidence=not_experimental
               /rpt_family="MIR"
               5062. .5157
               /note="homology = 67.70%, score = 32, counts = 48"
               /evidence=not_experimental
               /rpt_type=tandem
               /rpt_unit=tc
               5649. .5691
               /evidence=not_experimental
               /rpt_family="MIR"
               6131. .13747
               /gene="GATA-1"
               join(6131. .6168,10614. .10852,11367. .11744,11846. .11991,
               12695. .12820,13316. .13747)
               /gene="GATA-1"
               /product="GATA-binding protein 1"
               /note="mRNA acc. no. M30601"
               6131. .6168
               /gene="GATA-1"
               /number=1
               8548. .8857
               /evidence=not_experimental
               /rpt_family="AluSx"
               9694. .9979
               /evidence=not_experimental
               /rpt_family="AluJb"
               10614. .10852
               /gene="GATA-1"
               /number=2
               join(10633. .10852,11367. .11744,11846. .11991,12695. .12820,
               13316. .13687)
               /gene="GATA-1"
               /note="globin transcription factor 1; GF-1"
               /codon_start=1
               /product="GATA-binding protein 1"
               /protein_id="AAF06806.1"
               /db_xref="GI:6289081"
               /translation="MEFFGLGSLGTSELPQVDPALVSSTPESGVFFPSGEGDLAA
               ASSTAPSTATAAAALAYRDAAEYRHSVFQVYPLNLMGEGIPGSGPYAGWAYGKVTG
               LYPASTVCTREDSPQAVEDLDGKSTSFLETKRLSPDLLTGLPALPSLPVPN
               SAYGGPDSFTSPFTGSPPLNSAAYSKPLRGTLPDPCAEARECVNCGATAPLWRRD
               RTGHVLCNAGLYHKMNGONRPLIRPKRLIYSKAGTQCTNCOTTTTLWRNNSGD
               PNCAGGLYKLVHQNRLTRKKGIOIRNRKASGKRGKSSLGGAAGGAGGPF
               MVVAGGSGNCGEAVAGSLTGLPGTGAHLIYQGLGVPVLVSGPVSHLMPFPGPLLGSPTG
               SPTGPMPTTSTTVVAPLSS"
               11367. .11744
               /gene="GATA-1"
               /number=3
               11846. .11991
               /gene="GATA-1"
               /number=4
               complement(12330. .12477)
               /evidence=not_experimental
               /rpt_family="MER5B"
               12695. .12820
               /gene="GATA-1"
               /number=5
               13316. .13747
               /gene="GATA-1"
               /number=6
               14173. .14226
               /note="homology = 98.10%, score = 25, counts = 2"
               /evidence=not_experimental
               /rpt_type=tandem
               /rpt_unit=ggcatgacgctggggacaccttttagac
               14290. .14375
               /note="homology = 62.80%, score = 20, counts = 43"
               /evidence=not_experimental
               /rpt_type=tandem
               /rpt_unit=gt
               complement(14876. .14972)
               /evidence=not_experimental
               /rpt_family="MIR"

```

```
repeat_region complement(14899..15178)
/evidence-not_experimental
/rpt_family="MIR"
16160..16259
/note="homology = 74.00%, score = 23, counts = 4"
/evidence-not_experimental
/rpt_type=tandem
/rpt_unit=atacaggaacccctgtctcaccat
16338..16674
/evidence-not_experimental
/rpt_family="MERLB"
16780..16836
/note="homology = 86.00%, score = 22, counts = 3"
/evidence-not_experimental
/rpt_type=tandem
/rpt_unit=ttatttacaggaatctctg
16853..17048
/note="homology = 75.00%, score = 24, counts = 4"
/evidence-not_experimental
/rpt_type=tandem
/rpt_unit=ctactgtctgaacccacaaggaa
complement(17629..17940)
/evidence-not_experimental
/rpt_family="MER33"
18028..18353
/evidence-not_experimental
/rpt_family="MLTII"
complement(18388..18668)
/evidence-not_experimental
/rpt_family="AluSx"
complement(18674..18977)
/evidence-not_experimental
/rpt_family="AluSg"
19143..19448
/evidence-not_experimental
/rpt_family="AluJb"
19750..19923
/evidence-not_experimental
/rpt_family="MER58C"
complement(20505..20587)
/evidence-not_experimental
/rpt_family="MIR"
20598..20658
/evidence-not_experimental

Query Match 38.3%; Score 375; DB 9; Length 113853;
Best Local Similarity 99.6%; Pred. No. 2.2e-170;
Matches 475; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 170 AGGAGCCCGCGGCTAAGCGACGCTCGCCCTTGGCTGAGACTGGCGCTCCTCGCC 229
|
|
|
Db 51869 AGGAGCCCGCGGCTAAGCGACGCTCGCCCTTGGCTGAGACTGGCGCTCCTCGCC 51810

QY 230 GCTTCGGCGGTCTAGTCCCGGAGGTGAGCGCGCGGGGGGGTCCAGAGTGGCGGG 289
|
|
|
Db 51809 GCTTCGGCGGTCTAGTCCCGGAGGTGAGCGCGCGGGGGGGTCCAGAGTGGCGGG 51750

QY 290 CGCTGGCGCATCTCTGGAGGCCGGAAGCTCAGAGCGGGCGCGCGCGCGGAGGAGG 349
|
|
|
Db 51749 CGCTGGCGCATCTCTGGAGGCCGGAAGCTCAGAGCGGGCGCGCGCGGAGGAGG 51690

QY 350 CTGAGGATCAGAGCGCGGTCTCTGGCGAGCTGCTGCGCGTCTGGGGGGCCCCCGCA 409
|
|
|
Db 51689 CTGAGGATCAGAGCGCGGTCTCTGGCGAGCTGCTGCGCGTCTGGGGGGCCCCCGCA 51630

QY 410 ACTCTGATCCGGCTCTGGCTTGACACGACGCCCGGCGGCTGCGAGCTCGCTC 469
|
|
|
Db 51629 ACTCTGATCCGGCTCTGGCTTGACACGACGCCCGGCGGCTGCGAGCTCGCTC 51570

QY 470 GCGCTCTGCTCCGGCGCGGCTTGAACCTCTGCGGCTAGAGCCAGCTTGTCCCGCGC 529
|
|
|
Db 51569 GCGCTCTGCTCCGGCGCGGCTTGAACCTCTGCGGCTAGAGCCAGCTTGTCCCGCGC 51510
```

```
QY 530 CCGTCCCGCGCGCGCTCCGACCCCGCGCGCGCGTCTACGACAGCGCCCGCGGCGCC 589
|
|
|
Db 51509 CCGTCCCGCGCGCGCTCCGACCCCGCGCGCGCGTCTACGACAGCGCCCGCGGCGCC 51450

QY 590 CGGATGCTGAGGAGCGAGCGACGACACCGGACGCTGGACCGCGAGCTGTGAGGT 646
|
|
|
Db 51449 CGGATGCTGAGGAGCGAGCGACGACACCGGACGCTGGACCGCGAGCTGTGAGGT 51393

RESULT 6
AX336215/c 301 bp DNA linear PAT 09-JAN-2002
LOCUS Sequence 6724 from Patent WO0194629.
DEFINITION AX336215
ACCESSION AX336215
VERSION AX336215.1 GI:18126934
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
Horrigan, S., Soppet, D.R. and Weaver, Z.
TITLE Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A 6724 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES Location/Qualifiers
source 1..301
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 41 a 79 c 131 g 48 t 2 others
ORIGIN

Query Match 23.1%; Score 226; DB 6; Length 301;
Best Local Similarity 99.6%; Pred. No. 3.6e-98;
Matches 276; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 703 CGCGCGCTCCGCGTCCGCGGACGACGATGTGGCTGTGAGCTGCCCGCGGCGGTG 762
|
|
|
Db 301 CGCGCGCTCCGCGTCCGCGGACGACGATGTGGCTGTGAGCTGCCCGCGGCGGTG 242

QY 763 CTGGGGCGCTGTGCTGTGTAAGCGCTAGAGACCCCGCGCGCGCGCTGCGAGGC 822
|
|
|
Db 241 CTGGGGCGCTGTGCTGTGTAAGCGCTAGAGACCCCGCGCGCGCTGCGAGGC 182

QY 823 CGCTCTTGCCACCTGAGCACTGCCGGATGCCGTCACCTGGGACCCGAGAGTGGCC 882
|
|
|
Db 181 CGCTCTTGCCACCTGAGCACTGCCGGATGCCGTCACCTGGGACCCGAGAGTGGCC 122

QY 883 CGGCATCCGCGCACCGAGGACTGCTCCCGCAGCAGCTCCAGAGCACTTACCCCGGCC 942
|
|
|
Db 121 CGGCATCCGCGCACCGAGGACTGCTCCCGCAGCAGCTCCAGAGCACTTACCCCGGCC 62

QY 943 AGCCAGCGCTCTACCCGAGGATCCCTACCCCGTGGC 979
|
|
|
Db 61 AGCCAGCGCTCTACCCGAGGATCCCTACCCCGTGGC 25

RESULT 7
AX409054/c 301 bp DNA linear PAT 14-JUN-2002
LOCUS Sequence 1701 from Patent WO0229103.
DEFINITION AX409054
ACCESSION AX409054
VERSION AX409054.1 GI:21441759
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J.G.
TITLE Gene expression profiles in liver cancer
```

```

/clone_lib="NCI_CGAP_Co24"
/lab_host="DH10B"
/notes="Vector: CMV-CDS0002"

```

```
CDS      20..796
/codon_start=1
/product="Similar to granin-like neuroendocrine peptide precursor"
/protein_id="AAH12263.1"
/db_xref="GI:15128670"
/translation="MAGSPLLCGPRAGGVGILVLLGLLRLPPTLSARPVKEPRSL
AASAPLVETSTPLRLRAVRGEAGAVQELARALAHLLAEQRERARAEADQQ
ARVLAQLRWGSPRASDPPLAPDDPDPAQAOLARALLRLDPAALAAQLVPAPAA
APRPVPVYDDGTPGPDVEDAGDETDPDPELLRYLLGRLLTGSSEPEAAPRRLLR
SVDQDLGPEVPENVLGALLRVKRLNPSQAPARRLLPP"

BASE COUNT 173 a 353 c 304 g 145 t
ORIGIN

Query Match 3.3%; Score 32; DB 10; Length 975;
Best Local Similarity 100.0%; Pred. No. 0.00023;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 TGGGGCAGCATGGCGGGTGCCTGCTCTG 80
|||||
Db 11 TGGGGCAGCATGGCGGGTGCCTGCTCTG 42

RESULT 10
AF293356 1025 bp mRNA linear ROD 31-AUG-2000
LOCUS Mus musculus IAA mRNA, complete cds.
DEFINITION AF293356
ACCESSION AF293356
VERSION AF293356.1 GI:9954415
KEYWORDS
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 1025)
AUTHORS Donadel,G., Marinos,N., DeSilva,M.G., Lu,J., Notkins,A.L. and Lan,M.S.
TITLE Molecular cloning and characterization of a highly basic protein,
IA-4, expressed in pancreatic islets and brain
JOURNAL Neuroendocrinology 67 (3), 190-196 (1998)
MEDLINE 98292390
PUBMED 9630436
REFERENCE 2 (bases 1 to 1025)
AUTHORS Donadel,G., Notkins,A.L. and Lan,M.S.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-2000) Laboratory of Oral Medicine, National
Institute of Dental Research, 9000 Rockville Pike, Building 30 Room
124, Bethesda, MD 20892, USA
FEATURES
source
1..1025
/organism="Mus musculus"
/db_xref="taxon:10090"
/cell_line="insulinoma b-TC-1"
94..557
/notes="highly basic protein; highly expressed in
pancreatic islets and neuronal cells"
/codon_start=1
/product="IA4"
/protein_id="AAG09048.1"
/db_xref="GI:9954416"
/translation="MAGSPLLCGPRAGGVGILVLLGLLRLPPTLSARPVKEPRSL
AASAPLVETSTPLRLRAVRGEAGAVGAGAGAGAGAGAGGDRNRVLRRLRIS
RMSWRSSCAGPALVRPRTPLAPDDPDPAQAOLARALLRLDPAALAAQLVPAPAA
PAPRPVPVYDDGPHWRRCRRRDW"

BASE COUNT 156 a 382 c 330 g 157 t
ORIGIN

Query Match 3.3%; Score 32; DB 10; Length 1025;
Best Local Similarity 100.0%; Pred. No. 0.00023;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 TGGGGCAGCATGGCGGGTGCCTGCTCTG 80
|||||
Db 11 TGGGGCAGCATGGCGGGTGCCTGCTCTG 42

RESULT 11
AF181560 1027 bp mRNA linear ROD 27-JAN-2000
LOCUS Mus musculus proSAAS gene, complete cds.
DEFINITION AF181560
ACCESSION AF181560
VERSION AF181560.1 GI:6653208
KEYWORDS
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1027)
AUTHORS Fricker,L., McKinzie,A.A., Sun,J., Curran,E., Qian,Y., Yan,L.,
Patterson,S.D., Courchesne,P.L., Richards,B., Levin,N., Mzhavia,N.,
Devi,L.A. and Douglass,J.
TITLE Identification and characterization of proSAAS, a granin-like
neuroendocrine peptide precursor that inhibits prohormone
processing
JOURNAL J. Neurosci. 20 (2), 639-648 (2000)
MEDLINE 20098938
PUBMED 10632593
REFERENCE 2 (bases 1 to 1027)
AUTHORS Fricker,L., McKinzie,A.A., Sun,J., Curran,E., Qian,Y., Yan,L.,
Patterson,S.D., Courchesne,P.L., Richards,B., Levin,N., Mzhavia,N.,
Devi,L.A. and Douglass,J.
TITLE Direct Submission
JOURNAL Submitted (30-AUG-1999) Department of Molecular Pharmacology,
Albert Einstein College of Medicine, 1300 Morris Park Avenue,
Bronx, NY 10461, USA
FEATURES
source
1..1027
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="FLMSAAS"
1..109
/notes="inhibits prohormone processing"
/function="inhibits prohormone processing"
/codon_start=1
/product="proSAAS"
/db_xref="GI:6653209"
/translation="MAGSPLLCGPRAGGVGILVLLGLLRLPPTLSARPVKEPRSL
AASAPLVETSTPLRLRAVRGEAGAVQELARALAHLLAEQRERARAEADQQ
ARVLAQLRWGSPRASDPPLAPDDPDPAQAOLARALLRLDPAALAAQLVPAPAA
APRPVPVYDDGTPGPDVEDAGDETDPDPELLRYLLGRLLTGSSEPEAAPRRLLR
SVDQDLGPEVPENVLGALLRVKRLNPSQAPARRLLPP"

sig_peptide 110..208
3'UTR 887..1027
polyA_signal 999..1004
polyA_site 1017
BASE COUNT 149 a 386 c 332 g 160 t
ORIGIN

Query Match 3.3%; Score 32; DB 10; Length 1027;
Best Local Similarity 100.0%; Pred. No. 0.00023;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 TGGGGCAGCATGGCGGGTGCCTGCTCTG 80
|||||
Db 101 TGGGGCAGCATGGCGGGTGCCTGCTCTG 132

RESULT 12
AC098339 62396 bp DNA linear HTG 24-AUG-2002
LOCUS Rattus norvegicus clone CH230-2A16, *** SEQUENCING IN PROGRESS ***
DEFINITION AC098339
ACCESSION AC098339
VERSION AC098339.3 GI:22474728
```

KEYWORDS
SOURCE
ORGANISM

HTG: HTGS-PHASE1.
Norway rat.
Rattus norvegicus
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Eumariota; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS

1 (bases 1 to 62396)
Murny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbala,J., Benton,J., Blmage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,
Honsi,F., Howard,S., Huber,J.J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jollivet,S., Joudah,S.,
Karissom,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mahoney,E., McLeod,M.P., Meador,M., Mel,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okwuonu,G.,
Oragunye,N., Oviado,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rivers,M., Rojas,A., Rojokhan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shookstari,N., Sisson,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,T., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczkyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

TITLE
JOURNAL

Unpublished
2 (bases 1 to 62396)
Worley,K.C.
Direct Submission
Submitted (23-OCV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS
JOURNAL

3 (bases 1 to 62396)
Worley,K.C.
Direct Submission
Submitted (24-AUG-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Aug 24, 2002 this sequence version replaced gi:21953722.

COMMENT

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: TULN
Center clone name: CH230-2A16
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye: 100% of reads

Assembly program: Phrap; version 0.990329
Consensus quality: 23320 bases at least Q40
Consensus quality: 26055 bases at least Q30
Consensus quality: 27202 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
consists of 38 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 1088: contig of 1088 bp in length
1089 1188: gap of unknown length
1189 2332: contig of 1144 bp in length
2333 2432: gap of unknown length
2433 3637: contig of 1205 bp in length
3638 3737: gap of unknown length
3738 4884: contig of 1147 bp in length
4885 4984: gap of unknown length
4985 6231: contig of 1247 bp in length
6232 6331: gap of unknown length
6332 7397: contig of 1066 bp in length
7398 7497: gap of unknown length
7498 8602: contig of 1105 bp in length
8603 8702: gap of unknown length
8703 9708: contig of 1006 bp in length
9709 9808: gap of unknown length
9809 11095: contig of 1287 bp in length
11096 11195: gap of unknown length
11196 12589: contig of 1394 bp in length
12590 13689: gap of unknown length
13689 13967: contig of 1178 bp in length
13968 15150: contig of 1183 bp in length
15151 15250: gap of unknown length
15251 16391: contig of 1141 bp in length
16392 16491: gap of unknown length
16492 17528: contig of 1037 bp in length
17529 17628: gap of unknown length
17629 18906: contig of 1278 bp in length
18907 19006: gap of unknown length
19007 20152: contig of 1146 bp in length
20153 20252: gap of unknown length
20253 21292: contig of 1040 bp in length
21293 21392: gap of unknown length
21393 22700: contig of 1308 bp in length
22701 22800: gap of unknown length
22801 24706: contig of 1906 bp in length
24707 24806: gap of unknown length
24807 26388: contig of 1582 bp in length
26389 26488: gap of unknown length
26489 27637: contig of 1149 bp in length
27638 27737: gap of unknown length
27738 29102: contig of 1365 bp in length
29103 29202: gap of unknown length
29203 30352: contig of 1150 bp in length
30353 30452: gap of unknown length
30453 31720: contig of 1268 bp in length
31721 31820: gap of unknown length
31821 33903: contig of 2083 bp in length
33904 34003: gap of unknown length
34004 35759: contig of 1756 bp in length
35760 35859: gap of unknown length
35860 37240: contig of 1381 bp in length
37241 37340: gap of unknown length
37341 39222: contig of 1882 bp in length
39223 39322: gap of unknown length
39323 40923: contig of 1601 bp in length
40924 41023: gap of unknown length

```

* 41024 42882: contig of 1859 bp in length
* 42883 42982: gap of unknown length
* 42983 44647: contig of 1665 bp in length
* 44648 44747: gap of unknown length
* 44748 46546: contig of 1799 bp in length
* 46547 46646: gap of unknown length
* 46647 48461: contig of 1815 bp in length
* 48462 48561: gap of unknown length
* 48562 51223: contig of 2662 bp in length
* 51224 51323: gap of unknown length
* 51324 53891: contig of 2568 bp in length
* 53892 53991: gap of unknown length
* 53992 56351: contig of 2360 bp in length
* 56352 56451: gap of unknown length
* 56452 58948: contig of 2497 bp in length
* 58949 59048: gap of unknown length
* 59049 62396: contig of 3348 bp in length.
FEATURES             Location/Qualifiers
     source           1..62396
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="CH230-2A16"
BASE COUNT      15738 a 13739 c 13230 g 15967 t 3722 others
ORIGIN

Query Match      3.3%; Score 32; DB 2; Length 62396;
Best Local Similarity 100.0%; Pred. No. 0.00017;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 268 GCGGTGAGGAGTGGCGGCGGCGCTGGCGCA 299
      |||||||||||||||||||||||||||||||
Db 54260 GCGGTGAGGAGTGGCGGCGGCGCTGGCGCA 54291

RESULT 13
AC098473
LOCUS      Rattus norvegicus clone CH230-2A23, *** SEQUENCING IN PROGRESS ***,
DEFINITION      36 unordered pieces.
ACCESSION      AC098473.3 GI:21953422
VERSION        AC098473
KEYWORDS       HTG; HTGS_PHASE1.
SOURCE         Norway rat.
ORGANISM       Rattus norvegicus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
               Rattus.
REFERENCE      1 (bases 1 to 77876)
AUTHORS       Muzny,D.M., Adams,C., Adlo-Oduola,B., Ali-osman,F.R., Allen,C.,
               Alsbrooks,S.L., Anaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
               Barbara,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
               Bouck,J., Bowler,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
               Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
               Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
               Chen,G., Chen,K., Chen,Z., Chowdhry,I., Christopoulos,C.,
               Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
               Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
               Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
               Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
               Falls,T., Ferraguto,D., Flagg,N., Ford,J.J., Foster,P., Frantz,P.,
               Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
               Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
               Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
               Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
               Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
               Jacobson,B., Jia,L., Johnson,R., Jollivet,S., Joudah,S.,
               Karlsson,E., Kelly,S., Khan,U., King,D., Korvah,J., Kovar,C.,
               Kratovic,J., Kureishi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
               Li,J., Li,X., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Lousseged,H.,
               Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
               Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
               Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,

```

```

Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I.,
Sodergren,E., Sonaika,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tatney,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villaion,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Woodden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 77876)
Worley,K.C.
Direct Submission
Submitted (23-OCT-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 77876)
Worley,K.C.
Direct Submission
Submitted (31-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 24, 2002 this sequence version replaced gl:20976551.
----- Genome Center
Center: Baylor College of Medicine
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: TUVZ
Center clone name: CH230-2A23
----- Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 42926 bases at least Q40
Consensus quality: 42726 bases at least Q30
Consensus quality: 50269 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 36 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1471: contig of 1471 bp in length
* 1472 1571: gap of unknown length
* 1572 3224: contig of 1653 bp in length
* 3225 3324: gap of unknown length
* 3325 4401: contig of 1077 bp in length
* 4402 4501: gap of unknown length
* 4502 5530: contig of 1029 bp in length
* 5531 5630: gap of unknown length
* 5631 6686: contig of 1056 bp in length
* 6687 6786: gap of unknown length
* 6787 8420: contig of 1634 bp in length
* 8421 8521: gap of unknown length
* 8522 9651: contig of 1130 bp in length
* 9652 10758: contig of 1008 bp in length
* 10759 12472: contig of 1614 bp in length

```



```

* 12473 12572: gap of unknown length
* 12573 13767: contig of 1195 bp in length
* 13768 13867: gap of unknown length
* 13868 15438: contig of 1571 bp in length
* 15439 15538: gap of unknown length
* 15539 16997: contig of 1459 bp in length
* 16998 17097: gap of unknown length
* 17098 18756: contig of 1659 bp in length
* 18757 18856: gap of unknown length
* 18857 20149: contig of 1293 bp in length
* 20150 20249: gap of unknown length
* 20250 22457: contig of 2208 bp in length
* 22458 22557: gap of unknown length
* 22558 24293: contig of 1736 bp in length
* 24294 24393: gap of unknown length
* 24394 26264: contig of 1871 bp in length
* 26265 26364: gap of unknown length
* 26365 28263: contig of 1899 bp in length
* 28264 28364: gap of unknown length
* 28364 30529: contig of 2166 bp in length
* 30530 30629: gap of unknown length
* 30630 33390: contig of 2761 bp in length
* 33391 33490: gap of unknown length
* 33491 35353: contig of 1863 bp in length
* 35354 35453: gap of unknown length
* 35454 37565: contig of 2112 bp in length
* 37566 37665: gap of unknown length
* 37666 39106: contig of 1441 bp in length
* 39107 39206: gap of unknown length
* 39207 41043: contig of 1837 bp in length
* 41044 41143: gap of unknown length
* 41144 42556: contig of 1313 bp in length
* 42557 42556: gap of unknown length
* 42557 46801: contig of 4245 bp in length
* 46802 46901: gap of unknown length
* 46902 50115: contig of 3214 bp in length
* 50116 50215: gap of unknown length
* 50216 52326: contig of 2311 bp in length
* 52327 52626: gap of unknown length
* 52627 55467: contig of 2841 bp in length
* 55468 55567: gap of unknown length
* 55568 57414: contig of 1847 bp in length
* 57415 57514: gap of unknown length
* 57515 60035: contig of 2521 bp in length
* 60036 60135: gap of unknown length
* 60136 63450: contig of 3315 bp in length
* 63451 66350: gap of unknown length
* 63551 66353: contig of 3083 bp in length
* 66354 66733: gap of unknown length
* 66734 70948: contig of 4215 bp in length
* 70949 71048: gap of unknown length
* 71049 74326: contig of 3278 bp in length
* 74327 74426: gap of unknown length
* 74427 77876: contig of 3450 bp in length.

FEATURES
    Location/Qualifiers
        1..77876
            /organism="Rattus norvegicus"
            /db_xref="taxon:10116"
            /clone="CH230-2J23"
BASE COUNT 18873 a 17600 c 17320 g 19098 t 4985 others
ORIGIN

```

```

Query Match      3.3%; Score 32; DB 2; Length 77876;
Best Local Similarity 100.0%; Pred. No. 0.00017;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 268 GCGGTGAGAGCTGCGGGGCGCTGGCGCA 299
Db 38507 GCGGTGAGAGCTGCGGGGCGCTGGCGCA 38538

```

```

RESULT 14
AC094563/c

```

```

LOCUS
DEFINITION
AC094563
Rattus norvegicus clone CH230-4K5, *** SEQUENCING IN PROGRESS ***,
53 unordered pieces.
AC094563
VERSION
AC094563.3 GI:21717879
HTG: HTGS_PHASE1.
KEYWORDS
Norway rat.
SOURCE
Rattus norvegicus
Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 111948)
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
Alsbrooks, S.L., Amaratunga, H.C., Are, J.R., Ayelle, M., Banks, T.,
Barbaria, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D.,
Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Carton, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
Fallis, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, F., Johnson, R., Jolivet, S., Joudah, S.,
Karlssoon, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseghe, H.,
Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
Massey, E., Mawhinney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogum, M., Okwuonu, G.,
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
Rives, M., Rojas, A., Rojubokan, I., Roife, M., Ruiz, S., Savary, G.,
Scherrer, S., Scott, G., Shen, H., Shoshitari, N., Sisson, I.,
Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 111948)
Worley, K.C.
Direct Submission
Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 111948)
Worley, K.C.
Direct Submission
Submitted (11-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 10, 2002 this sequence version replaced gi:17941324.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project information

```

```

COMMENT

```

Center project name: GAXL
Center clone name: CH230-4K5
----- Summary Statistics -----
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 54807 bases at least Q40
Consensus quality: 60236 bases at least Q30
Consensus quality: 64720 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 53 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1001: contig of 1001 bp in length
* 1002: gap of unknown length
* 1101: contig of 1231 bp in length
* 2332: gap of unknown length
* 2333: gap of unknown length
* 2433: contig of 1053 bp in length
* 3485: gap of unknown length
* 3586: contig of 1229 bp in length
* 4814: gap of unknown length
* 4914: contig of 1240 bp in length
* 6154: contig of unknown length
* 6254: gap of unknown length
* 7917: contig of 1663 bp in length
* 8017: gap of unknown length
* 9057: contig of 1040 bp in length
* 9157: gap of unknown length
* 10364: contig of 1207 bp in length
* 10365: gap of unknown length
* 10465: contig of 1433 bp in length
* 11897: gap of unknown length
* 11898: contig of 1266 bp in length
* 13263: gap of unknown length
* 13264: contig of 1181 bp in length
* 13364: gap of unknown length
* 14544: contig of 1420 bp in length
* 14644: gap of unknown length
* 16064: contig of 1621 bp in length
* 16164: gap of unknown length
* 16165: contig of 1238 bp in length
* 17786: gap of unknown length
* 17886: contig of 1238 bp in length
* 18986: contig of 1408 bp in length
* 19124: gap of unknown length
* 19224: contig of 1408 bp in length
* 20632: gap of unknown length
* 20732: contig of 1483 bp in length
* 22215: gap of unknown length
* 22314: contig of 1435 bp in length
* 23749: gap of unknown length
* 23849: gap of unknown length
* 25423: contig of 1574 bp in length
* 25523: gap of unknown length
* 26850: contig of 1327 bp in length
* 26950: gap of unknown length
* 28487: contig of 1537 bp in length
* 28488: gap of unknown length
* 28588: contig of 1249 bp in length
* 29936: gap of unknown length
* 31672: contig of 1736 bp in length
* 31772: gap of unknown length
* 31773: contig of 1389 bp in length
* 33161: gap of unknown length
* 33261: contig of 1176 bp in length
* 34437: gap of unknown length
* 34537: contig of 2215 bp in length
* 34538: gap of unknown length
* 36753: gap of unknown length
* 37909: contig of 1057 bp in length
* 38009: gap of unknown length
* 39170: contig of 1161 bp in length

39171 39270: gap of unknown length
* 39271 41040: contig of 1770 bp in length
* 41041 41140: gap of unknown length
* 41141 42441: contig of 1301 bp in length
* 42442 42541: gap of unknown length
* 42542 43610: contig of 1069 bp in length
* 43611 43710: gap of unknown length
* 43711 45019: contig of 1309 bp in length
* 45020 45119: gap of unknown length
* 45120 46358: contig of 1239 bp in length
* 46359 46458: gap of unknown length
* 46459 48595: contig of 2137 bp in length
* 48596 50326: gap of unknown length
* 50327 50426: gap of unknown length
* 50427 51957: contig of 1531 bp in length
* 51958 52057: gap of unknown length
* 52058 53961: contig of 1904 bp in length
* 53962 54061: gap of unknown length
* 54062 57027: contig of 2966 bp in length
* 57028 57127: gap of unknown length
* 57128 58760: contig of 1633 bp in length
* 58761 58860: gap of unknown length
* 58861 61388: contig of 2528 bp in length
* 61389 63296: contig of 1808 bp in length
* 63297 63396: gap of unknown length
* 63397 65907: contig of 2511 bp in length
* 65908 66007: gap of unknown length
* 66008 69562: contig of 3555 bp in length
* 69563 72572: contig of 2910 bp in length
* 72573 72672: gap of unknown length
* 72673 75269: contig of 2597 bp in length
* 75270 75369: gap of unknown length
* 75370 78347: contig of 2978 bp in length
* 78348 82487: gap of unknown length
* 82488 82587: gap of unknown length
* 82588 84648: contig of 2061 bp in length
* 84649 87601: contig of 2853 bp in length
* 87602 87701: gap of unknown length
* 87702 91381: contig of 3679 bp in length
* 91382 91481: gap of unknown length
* 91482 95059: contig of 3479 bp in length
* 95060 95060: gap of unknown length
* 95061 99432: contig of 4373 bp in length
* 99433 99532: gap of unknown length
* 99533 105198: contig of 5666 bp in length
* 105199 111948: gap of unknown length
* 105299 111948: contig of 6650 bp in length.

FEATURES
source
Location/Qualifiers
1. 111948

Query Match 3.3%; Score 32; DB 2; Length 111948;
Best Local Similarity 100.0%; Pred. No. 0.00017;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 TGGGGCAGCATGGCGGGTGGCGCTGCTCTG 80
|||||
Db 109507 TGGGGCAGCATGGCGGGTGGCGCTGCTCTG 109476

RESULT 15

AL670169
LOCUS Mouse DNA sequence from clone RP23-198C2 on chromosome X, complete
DEFINITION sequence.
AL670169
ACCESSION AL670169.8 GI:21213366
VERSION AL670169.8
KEYWORDS house mouse.
HYG.

ORGANISM Mus musculus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 151694)
 Direct Submission
 Submitted (23-MAY-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On May 25, 2002 this sequence version replaced gi:21104114.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em: EMBL; Sw:,
 SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
 database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-198C2 is
 from the RPCI-23 Mouse PAC Library
 constructed by the group of Pieter de Jong.
 For further details see <http://www.chori.org/bacpac/home.htm>
 VECTOR: PBACe3.6.

FEATURES
 source
 1..151694
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="X"
 /clone="RP23-198C2"
 /clone_lib="RPCI-23"
 BASE COUNT 39029 a 35150 c 35380 g 42135 t
 ORIGIN

Query Match 3.3%; Score 32; DB 10; Length 151694;
 Best Local Similarity 100.0%; Pred. No. 0.00016;
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 49 TGGGCGAGCATGGCGGGTGGCGGCTGCTCTG 80
 |||||
 Db 21997 TGGGCGAGCATGGCGGGTGGCGGCTGCTCTG 22028

Search completed: April 12, 2003, 20:24:33
 Job time : 3039 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2003, 14:26:00 ; Search time 45 Seconds
(without alignments)
555.443 Million cell updates/sec

Title: US-09-803-589-6
Perfect score: 260
Sequence: 1 MAGSPLLWGPRAGGVGLLV.....RVKRLETPAPQVPARLLPP 260

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 10

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	%	
No.	Score	Match Length DB ID	Description

No matches found

Search completed: April 4, 2003, 14:29:59
Job time : 46 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2003, 14:18:45 ; Search time 25 Seconds
(without alignments)
431.354 Million cell updates/sec

Title: US-09-803-589-6
Perfect score: 260
Sequence: 1 MACSPLLWGPRAGGVCLLVL.....RVKRLETPAPQVPARRLLPP 260

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0
Searched: 112892 seqs, 41476328 residues

Word size : 10
Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

No matches found

Search completed: April 4, 2003, 14:27:08
Job time : 25 secs


```
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 63 VPRGEAGAGVQLARALAHLLAEERQERARAEQAEDQARVLAQLLR 111
      |||
Db 63 VPRGEAGAGVQLARALAHLLAEERQERARAEQAEDQARVLAQLLR 111

RESULT 3
Q9QXU9 PRELIMINARY; PRT; 260 AA.
AC Q9QXU9;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PROSAA precursor.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20098938; PubMed=10632593;
RA Fricker L., McKinzie A.A., Sun J., Curran E., Qian Y., Van L.,
RA Patterson S.D., Courchesne P.L., Richards B., Levin N., Mzhavia N.,
RA Devi L.A., Douglass J.;
RT "Identification and characterization of prosaas, a granin-like
RT neuroendocrine peptide precursor that inhibits prohormone
RT processing.";
RL J. Neurosci. 20:639-648(2000).
DR EMBL; AF181561; AAF22642.1; -.
KW SIGNAL.
FT SIGNAL.
SQ SEQUENCE 260 AA; 27414 MW; C84F688BDEB5313E CRC64;

Query Match 18.8%; Score 49; DB 11; Length 260;
Best Local Similarity 100.0%; Pred. No. 1.4e-35;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 63 VPRGEAGAGVQLARALAHLLAEERQERARAEQAEDQARVLAQLLR 111
      |||
Db 63 VPRGEAGAGVQLARALAHLLAEERQERARAEQAEDQARVLAQLLR 111

RESULT 4
Q9QXV0 PRELIMINARY; PRT; 258 AA.
AC Q9QXV0;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE PROSAA precursor.
GN PCSK1N OR SAAS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20098938; PubMed=10632593;
RA Fricker L., McKinzie A.A., Sun J., Curran E., Qian Y., Van L.,
RA Patterson S.D., Courchesne P.L., Richards B., Levin N., Mzhavia N.,
RA Devi L.A., Douglass J.;
RT "Identification and characterization of prosaas, a granin-like
RT neuroendocrine peptide precursor that inhibits prohormone
RT processing.";
RL J. Neurosci. 20:639-648(2000).
DR EMBL; AF181560; AAF22641.1; -.
MCD; MGI:1353431; Pcskn.
DR InterPro; IPR002965; P_rich_extensn.
DR PRINTS; PR01217; PRICHEXTENS.
KW SIGNAL.
FT SIGNAL.
SQ SEQUENCE 258 AA; 27285 MW; 07452A460E868CF6 CRC64;
```

```
Query Match 12.7%; Score 33; DB 11; Length 258;
Best Local Similarity 100.0%; Pred. No. 2.8e-21;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 DDDPDAPAAQLARALLRARLDPAALAAQLVPAP 158
      |||
Db 126 DDDPDAPAAQLARALLRARLDPAALAAQLVPAP 158

RESULT 5
Q9ESU4 PRELIMINARY; PRT; 187 AA.
AC Q9ESU4;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE IA4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98292390; PubMed=9630436;
RA Donadel G., Marinos N., DeSilva M.G., Lu J., Notkins A.L., Lan M.S.;
RA "Molecular cloning and characterization of a highly basic protein, IA-
RT 4, expressed in pancreatic islets and brain.";
RL Neuroendocrinology 67:190-196(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Donadel G., Notkins A.L., Lan M.S.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF293356; AAG09048.1; -.
SQ SEQUENCE 187 AA; 19940 MW; 041F1F858591A9B0 CRC64;

Query Match 9.6%; Score 25; DB 11; Length 187;
Best Local Similarity 100.0%; Pred. No. 3e-14;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 DDDPDAPAAQLARALLRARLDPAAL 150
      |||
Db 127 DDDPDAPAAQLARALLRARLDPAAL 151

Search completed: April 4, 2003, 14:29:08
Job time : 33 secs
```

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2003, 14:26:46 ; Search time 28 Seconds
(without alignments)
273.213 Million cell updates/sec

Title: US-09-803-589-6
Perfect score: 260
Sequence: 1 MAGSPLWSPRAGGVGLLV.....RVKRLTPAPQVPARRLLPP 260

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 10

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description

No matches found

Search completed: April 4, 2003, 14:30:34
Job time : 28 secs


```

RESULT 1
US-09-803-589-6
: Sequence 6, Application US/09803589
: Patent No. US20020112251A1
: GENERAL INFORMATION
: APPLICANT: McCarthy, Sean A.
: APPLICANT: Holtzman, Douglas A.
: APPLICANT: Goodear, Andrew D.J.
: TITLE OF INVENTION: NOVEL GENES EN
: TITLE OF INVENTION: PROGNOSTIC, D

```

```
; SEQ ID NO 150
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (149)
; OTHER INFORMATION: Xaa equals stop translation
US-09-984-271-150
```

```
Query Match
Best Local Similarity 100.0%; Score 51; DB 9; Length 149;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 MAGSPLLWGPRAGGVLVLLGLLFRPPPPALCARPVKEPRGLSAASPPLA 51
|
Db 1 MAGSPLLWGPRAGGVLVLLGLLFRPPPPALCARPVKEPRGLSAASPPLA 51
|
```

RESULT 3

```
US-09-984-271-238
; Sequence 238, Application US/09984271
; Publication No. US20030040088A1
```

```
; GENERAL INFORMATION:
```

```
; APPLICANT: Rosen et al.
```

```
; TITLE OF INVENTION: 71 Human Secreted Proteins
```

```
; FILE REFERENCE: P2030P1
```

```
; CURRENT APPLICATION NUMBER: US/09/984,271
```

```
; CURRENT FILING DATE: 2001-10-29
```

```
; PRIOR APPLICATION NUMBER: 09/482,273
```

```
; PRIOR FILING DATE: 2000-01-13
```

```
; PRIOR APPLICATION NUMBER: PCT/US99/15849
```

```
; PRIOR FILING DATE: 1999-07-14
```

```
; PRIOR APPLICATION NUMBER: 60/092,921
```

```
; PRIOR FILING DATE: 1998-07-15
```

```
; PRIOR APPLICATION NUMBER: 60/092,922
```

```
; PRIOR FILING DATE: 1998-07-15
```

```
; PRIOR APPLICATION NUMBER: 60/092,956
```

```
; PRIOR FILING DATE: 1998-07-15
```

```
; NUMBER OF SEQ ID NOS: 267
```

```
; SOFTWARE: PatentIn Ver. 2.0
```

```
; SEQ ID NO 238
```

```
; LENGTH: 166
```

```
; TYPE: PRT
```

```
; ORGANISM: Homo sapiens
```

```
US-09-984-271-238
```

```
Query Match
Best Local Similarity 100.0%; Score 51; DB 9; Length 166;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 MAGSPLLWGPRAGGVLVLLGLLFRPPPPALCARPVKEPRGLSAASPPLA 51
|
Db 19 MAGSPLLWGPRAGGVLVLLGLLFRPPPPALCARPVKEPRGLSAASPPLA 69
|
```

RESULT 4

```
US-09-803-589-14
```

```
; Sequence 14, Application US/09803589
```

```
; Patent No. US20020112251A1
```

```
; GENERAL INFORMATION:
```

```
; APPLICANT: McCarthy, Sean A.
```

```
; APPLICANT: Holtzman, Douglas A.
```

```
; APPLICANT: Goodearl, Andrew D.J.
```

```
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
```

```
; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC AND OTHER
```

```
; TITLE OF INVENTION: USES
```

```
; FILE REFERENCE: 07334-325001
```

```
; CURRENT APPLICATION NUMBER: US/09/803,589
```

```
; CURRENT FILING DATE: 2001-03-09
```

```
; PRIOR APPLICATION NUMBER: US 09/128,709
```

```
; PRIOR FILING DATE: 1998-08-04
```

```
; PRIOR APPLICATION NUMBER: US 60/054,645
```

```
; PRIOR FILING DATE: 1997-08-04
; PRIOR APPLICATION NUMBER: US 09/130,491
; PRIOR FILING DATE: 1998-08-06
; PRIOR APPLICATION NUMBER: US 60/054,966
; PRIOR FILING DATE: 1997-08-06
; PRIOR APPLICATION NUMBER: US 60/058,108
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: US 09/388,280
; PRIOR FILING DATE: 1999-09-01
; PRIOR APPLICATION NUMBER: US 09/388,279
; PRIOR FILING DATE: 1999-09-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 175
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-803-589-14
```

```
Query Match
Best Local Similarity 100.0%; Score 49; DB 10; Length 175;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 63 VPRGEAGAGVQELARALAHLLLEAEERQERARAEAEQQAQARVLAQLLR 111
|
Db 63 VPRGEAGAGVQELARALAHLLLEAEERQERARAEAEQQAQARVLAQLLR 111
|
```

```
Search completed: April 4, 2003, 14:34:06
Job time : 16 secs
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 12, 2003, 17:01:47 ; Search time 1577 Seconds
(without alignments)
10054.142 Million cell updates/sec

Title: US-09-803-589-5

Perfect score: 979

Sequence: 1 gaattggcagcaggccagc.....gaggatccctaccctcgtgc 979

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- EST.*
- 1: em_estba.*
 - 2: em_esthum.*
 - 3: em_estin.*
 - 4: em_estnu.*
 - 5: em_estov.*
 - 6: em_estpl.*
 - 7: em_estro.*
 - 8: em_htc.*
 - 9: gb_est1.*
 - 10: gb_est2.*
 - 11: gb_htc.*
 - 12: gb_est3.*
 - 13: gb_est4.*
 - 14: gb_est5.*
 - 15: em_estfun.*
 - 16: em_estom.*
 - 17: gb_gss.*
 - 18: em_gss_hum.*
 - 19: em_gss_inv.*
 - 20: em_gss_pin.*
 - 21: em_gss_vrt.*
 - 22: em_gss_fun.*
 - 23: em_gss_mam.*
 - 24: em_gss_mus.*
 - 25: em_gss_other.*
 - 26: em_gss_pro.*
 - 27: em_gss_rod.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	757	77.3	839	13	BI546579
2	750.2	76.6	878	13	BI603988
c	739.6	75.5	951	13	BI489332
4	736.2	75.2	887	13	BI602351
5	735.6	75.1	764	13	BI669134
6	735.2	75.1	868	13	BI458350

7	727.4	74.3	908	13	BI668879
8	724.6	74.0	745	12	BG707647
9	721	73.6	861	12	BG709387
10	712.8	72.8	817	13	BI668984
c	697	71.2	990	12	BF966686
12	696.6	71.2	784	13	BI668730
13	691.8	70.7	707	14	BM715636
14	688.6	70.3	771	13	BI667137
15	686.6	70.1	763	12	BG699405
16	678.8	69.3	898	13	BI544506
17	677.6	69.2	843	13	BI599557
18	677.6	69.2	883	13	BI549788
19	676	69.1	877	13	BI666575
20	670.8	68.5	735	13	BI159920
21	666	68.0	870	12	BF969262
22	663.8	67.8	740	13	BI602131
23	661.2	67.5	690	12	BG716263
24	658.8	67.3	692	13	BI490760
c	657.2	67.1	713	14	BM674078
26	649.4	66.3	857	12	BG708547
27	642	65.6	924	12	BE798423
28	624.4	63.8	681	12	BG706595
29	621.8	63.5	953	13	BI161379
30	620.6	63.4	763	13	BI602706
31	611	62.4	676	13	BI548089
32	608.8	62.2	693	13	BI603114
33	593.4	60.6	600	13	BI603114
34	589.4	60.2	736	12	BG703846
35	587.6	60.0	661	12	BG714685
36	586.8	59.9	718	10	BE383172
c	584	59.7	593	13	BM504127
38	581.8	59.4	586	13	BM129592
39	580.8	59.3	944	11	AK002969
40	578.2	59.1	631	12	BG108317
41	569.6	58.2	577	13	BM129728
42	569.6	58.2	585	13	BM128297
43	568.4	58.1	633	12	BG703203
44	564.8	57.7	958	12	BF310234
45	563.6	57.6	575	13	BI547813

ALIGNMENTS

RESULT 1	BI546579	839 bp	mRNA	linear	EST 05-SEP-2001			
LOCUS	603191602F1 NIH_MGC_95	Homo sapiens	cdna	clone IMAGE:5262672 5',				
DEFINITION	mRNA sequence.							
ACCESSION	BI546579							
VERSION	BI546579.1	GI:15433891						
KEYWORDS	EST.							
SOURCE	human.							
ORGANISM	Homo sapiens							
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;							
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.							
TITLE	1 (bases 1 to 839)							
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/ .							
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapsb@mail.nih.gov Tissue Procurement: Miklos Palkovits, M.D., Ph.D. cdna Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN) cdna Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov Plate: L1AM11662 row: d column: 01 High quality sequence stop: 793. Location/Qualifiers							

QY 790 CTAGAGACCCC-GGCGCCCCAGGTGCCTGCACGCGCCCTCTTGGCCACCTGTAGCACTGCC 848
 Db 165 CTAGAGACCCCAGGCGCCCGCCAGTGCCTGCACGCGCCCTCTTGGCCACCTGTAGCACTGCC 106
 QY 849 CGGATCCCGTGCACCTGGAGCCAGAGTGCCTCCCGCCCATCCCGCCACAGGACTGCTC 908
 Db 105 CGGATCCCGTGA-CCTGGGACCCAGAGTGCCTCCCGCCCATCCCGCCACAGGACTGCTC 47
 QY 909 CCGCGCAGACGCTCCAGAGCACTTACCCCGCCGACCCAGCCCTCT 954
 Db 46 CCGCGCAGACGCTCCAGAGCAAC-TAAACCGCGCCGCGCCAGCCCTCT 2

RESULT 4
 BI602351
 LOCUS 603251134F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5302616 5',
 DEFINITION mRNA sequence.
 ACCESSION BI602351
 VERSION BI602351.1 GI:15495303
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (Bases 1 to 887)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
 Toshiyuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L2AM11766 row: d column: 09
 High quality sequence stop: 764.
 Location/Qualifiers
 1. 887
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5302616"
 /clone_lib="NIH_MGC_96"
 /tissue_type="hypothalamus"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: pBluescriptR (modified
 pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dr primed using primer 5'-TTTTTTTTTTTTTTVN-3',
 size-selected for average insert size 2.3 kb and
 normalized to ROT 5. This is a primary library enriched
 for full-length clones and constructed using the
 Cap-trapper method (Carninci, in preparation). Library
 constructed by M. Brownstein (NIH/NHGRI, National
 Institutes of Health). Note: this is a NIH_MGC Library."
 BASE COUNT 94 a 334 c 336 g 123 t
 ORIGIN

Query Match 75.2%; Score 736.2; DB 13; Length 887;
 Best Local Similarity 93.4%; Pred. No. 3.2e-113;
 Matches 819; Conservative 5; Mismatches 43; Indels 10; Gaps 5;

QY 15 GCCAGCAGTCCGCGCGYMGRRGCCCGCTCGCTGGGCGAGCATGGCGGGTCCCGCT 74
 Db 11 GCCAGCAGTCCGCGCGTCCGAGACCCCGCTCGCTGGGCGAGCATGGCGGGTCCCGCT 70
 QY 75 GCTCTGGGGCGCGGGCGCGGGCGCTCGCGCTTTTGTGCTGCTGCTGCTGCTGCTGCT 134
 Db 71 GCTCTGGGGCGCGGGCGCGGGCGCTCGCGCTTTTGTGCTGCTGCTGCTGCTGCTGCT 130

QY 135 TCGGGCGCGCGCGCGCGCTCTGCGGGGGCGCGGTAAAGGAGAGCCCGCGGCGCTAAGCGCAGC 194
 Db 131 TCGGGCGCGCGCGCGCGCTCTGCGGGCGCGGTAAAGGAGAGCCCGCGGCGCTAAGCGCAGC 190
 QY 195 GTCTCCGCGCTTGGCTGAGACTGGCGCTCTCGCGCGCTTCCGCGGTCAGTGCCTCCGAGS 254
 Db 191 GTCTCCGCGCTTGGCTGAGACTGGCGCTCTCGCGCGCTTCCGCGGTCAGTGCCTCCGAGS 250
 QY 255 TAGAGCGCGGGGGGGTGCAGAGAGTGGCGGGGGCGCTGGCGGCATCTCTCTGAGAGCCGA 314
 Db 251 TAGAGCGCGGGGGGGTGCAGAGAGTGGCGGGGGCGCTGGCGGCATCTCTCTGAGAGCCGA 310
 QY 315 ACCTCAGAGAGCGGGGGCGGGCGAGGCGAGGAGTGCAGATCAGCAGGCGGGGCTCT 374
 Db 311 ACCTCAGAGAGCGGGGGCGGGCGAGGCGAGGAGTGCAGATCAGCAGGCGGGGCTCT 370
 QY 375 GCGCAGAGTGTGCGCGTCTGGGGCGCGCCCGCCCAACTGTGATCGGCTCTCGGCTTGA 434
 Db 371 GCGCAGAGTGTGCGCGTCTGGGGCGCGCCCGCCCAACTGTGATCGGCTCTCGGCTTGA 430
 QY 435 CGACGACCCCGACGCGCCTGCAGCGAGCTCGCTCGCGCTCTGCTCCGCGCGCGCTTGA 494
 Db 431 CGACGACCCCGACGCGCCTGCAGCGAGCTCGCTCGCGCTCTGCTCCGCGCGCGCTTGA 490
 QY 495 CCCTCCGCGCTTAGCAGCGCCAGCTTGTCCCGCGCGCGCTCCCGCGCGCGCTCCGACC 554
 Db 491 CCCTCCGCGCTTAGCAGCGCCAGCTTGTCCCGCGCGCGCTCCCGCGCGCGCTCCGACC 550
 QY 555 CCGG--CCCGCGGTCTACGACGACGCGCGCGCGCGCGCGCTGAGGAGGAGCGGACG 613
 Db 551 CCGGCGCGCGGTCTACGACGACGCGCGCGCGCGCGCGCGCTGAGGAGGAGCGGACG 610
 QY 614 AGACACCCGACGTGACCCCGCGAGCTGTTGAGGTACTTGTCTGGG--ACGGATTCTTTTCGGG 671
 Db 611 AGACACCCGACGTGACCCCGCGAGCTGTTGAGGTACTTGTCTGGGAGCGGCTTTTCGGG 670
 QY 672 --AAGCGGGACTCCGAGGGGTGCGACGCGCGCGCGCGCTCCCGCGCGCGCTCCGACC 729
 Db 671 GAAGCGCGGACTCCGAGGGGTGCGACGCGCGCGCGCGCGCTCCG--CGTCCCGCGGAC 729
 QY 730 GATGTGGGCTGTAGCTGCGCCCTCAGGCGGTCTGGGGGCGCTGCTGC---GTGTGA 785
 Db 730 GATGTGGGCTGTAGCTGCGCCCTCAGGCGGTCTGGGGGCGCTGCTGGGGGCTGTGA 789
 QY 786 AGCGCTAGAGACCCCGCGCGCGCGCTGCTGACGCGCGCTCTGCGCGCTGAGCACT 845
 Db 790 AGCGCTAGAGACCCCGCGCGCGCGCTGCTGACGCGCGCTCTGCGCGCGCTGAGCACT 849
 QY 846 GCCCGATCCCGTGCACCTTGGGACCCAGAGTGC 882
 Db 850 TCGCGGATCCCGTGCACCTTGGGACCCAGAGTGC 886

RESULT 5
 BI669134
 LOCUS 603295564F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5314814 5',
 DEFINITION mRNA sequence.
 ACCESSION BI669134
 VERSION BI669134.1 GI:15583367
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (Bases 1 to 764)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki

[illegible]

Db	610	 GACATCCGACGTTGGACCCCGAGCTGTTGAGTACTTCTGGGACGGATCTTCGGGGAAG	669
Qy	675	CGCGAGCTCCGAGGGGTGGCAGCCCGCGCGCTCCGCGCTCCGCGCGACACGATGT	734
Db	670	CGCGAGCTCCGAGGGGTGGCAGCCCGCGCTCCGCGCTCCGCGCGACACGATGT	729
Qy	735	GGGCTCTGAGTGGCCCCCTGAGGGCGTGTGGGGC	770
Db	730	GGGCTCTGAGTGGCCCCCTGAGGGCGTGTGGGGC	764
RESULT 6			
BI458350			
LOCUS			
DEFINITION			
603198733F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5277993 5', mRNA sequence.			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
human.			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE			
1 (bases 1 to 868)			
NIH-MGC http://mgc.nci.nih.gov/ .			
National Institutes of Health, Mammalian Gene Collection (MGC)			
Unpublished (1999)			
COMMENT			
Contact: Robert Strausberg, Ph.D.			
Email: cgapbs-r@mail.nih.gov			
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.			
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki			
Toshiyuki and Piero Carninci (RIKEN)			
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)			
DNA Sequencing by: Incyte Genomics, Inc.			
Clone distribution: MGC clone distribution information can be			
found through the I.M.A.G.E. Consortium/LLNL at:			
http://image.llnl.gov			
Plate: L1AM11702 row: b column: 10			
High quality sequence stop: 838.			
FEATURES			
source			
1. 868			
Location/Qualifiers			
/organism="Homo sapiens"			
/db_xref="taxon:9606"			
/clone="IMAGE:5277993"			
/clone_lib="NIH_MGC_96"			
/tissue_type="hypothalamus"			
/lab_host="DH10B"			
/note="Organ: brain; Vector: pBluescriptR (modified			
pBluescript KS+); Site.1: BamHI; Site.2: SalI-XhoI (gtcgag			
pb); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',			
size-selected for average insert size 2.3 kb and			
normalized to Rot 5. This is a primary library enriched			
for full-length clones and constructed using the			
Cap-trapper method (Carninci, in preparation). Library			
constructed by M. Brownstein (NIH/NHGRI, National			
Institutes of Health). Note: this is a NIH_MGC Library."			
BASE COUNT			
ORIGIN			
94 a 324 c 325 g 125 t			
Query Match			
Best Local Similarity 75.1%; Score 735.2; DB 13; Length 868;			
Matches 822; Conservative 5; Mismatches 38; Indels 7; Gaps 6;			
Qy	13	AGGCAGCAGTCGCGCSGYMCGRRCCGCGCTCGCTGGGGCAGCATGGGGGTGCGCG	72
Db	1	AGCGAGGTCAGTCGCGCGCTCGGAGCCGCGCTAGCTGGGCGACATGGCGGTGCGCG	60
Qy	73	CTGCTCTGGGGCGCGCGGGCGGTGGCGCTTTTGGTCGTCTCTCGGCGCTG	132
Db	61	CTGCTCTGGGGCGCGCGGGCGGTGGCGCTTTTGGTCGTCTCTCGGCGCTG	120
Qy	133	TTTGGCGCGCGCGCGCGCTCTGCGCGCGCGCGCTAAGAGAGCCCGCGCTAAGCGCA	192

Db 121 TTTCGGCGCCCGCCGCTCTGCGGGCGGCGATAAAGAGGCCCGCGGCTTAAGCGCA 180
Qy 193 GCGTCTCCGCGCTTGGCTGAGACTGGCGCTCTCCGCGCTTCCGCGGTCACTGCGCCGA 252
Db 181 GCGTCTCCGCGCTTGGCTGAGACTGGCGCTCTCCGCGTTCGCGGTCACTGCGCCGA 240
Qy 253 GGTAGGCGCGCGCGCGGTGAGAGCTGGCGCGGGCGTGGCGATCTGCTGAGGCGC 312
Db 241 GGTAGGCGCGCGCGGTGAGAGCTGGCGCGGGCGTGGCGATCTGCTGAGGCGC 300
Qy 313 GAACGTCAGGAGCGCGCGCGGTGAGAGCTGGCGCGGGCGTGGCGATCTGCTGAGGCGC 372
Db 301 GAACGTCAGGAGCGCGCGGTGAGAGCTGGCGCGGGCGTGGCGATCTGCTGAGGCGC 360
Qy 373 CTGGCGCGAGCTCTGCGGCTTGGGGCGCGCCCGCACTCTGATCCGGCTCTGGGCTTG 432
Db 361 CTGGCGCGAGCTCTGCGGCTTGGGGCGCGCCCGCACTCTGATCCGGCTCTGGGCTTG 420
Qy 433 GACGACGCGCGCGCGGTGAGAGCTGGCGCGGGCGTGGCGATCTGCTGAGGCGC 492
Db 421 GACGACGCGCGCGGTGAGAGCTGGCGCGGGCGTGGCGATCTGCTGAGGCGC 480
Qy 493 GACGCTGCGCGCTAGAGCGCGAGC-TTGTCCCGCGCGCGTCCCGCGCGCGCGCTCGG 551
Db 481 GACGCTGCGCGCTAGAGCGCGAGC-TTGTCCCGCGCGCGTCCCGCGCGCGCGCTCGG 540
Qy 552 ACCCGCGCGCGGTCTACGAGCGCGCGCGCGCGATGCTGAGGAGCGAGCGGA 611
Db 541 ACCCGCGCGCGGTCTACGAGCGCGCGCGCGCGATGCTGAGGAGCGAGCGGA 600
Qy 612 CAGACACCGCGAGCGCGAGCTGTTGAGTACTTGTGCGGACGAGTCTTGTGCGGG 671
Db 601 CAGACACCGCGAGCGCGAGCTTGTGAGTACTTGTGCGGACGAGTCTTGTGCGGG 659
Qy 672 AAGCGCGAGCTCCGAGGCGGTGGAGCGCGCGCGCGCGTCCCGCGCGCGCGCGAGG- 730
Db 660 AAGCGCGAGCTCCGAGGCGGTGGAGCGCGCGCGCGCGTCCCGCGCGCGCGCGAGG 719
Qy 731 -ATGTGGGCTCTGAGCTCCCGCTGAGGCGGTGCTGGGGGCGCTGCTGCTGTGAAAGCG 789
Db 720 ATGTGGGCTCTGAGCTG-CCCGTGGAGCGGTGCTGGGGGCGCTGCTGCTGTGAAAGCG 778
Qy 790 CTAGAGACCGCGCGCGCGCGAGTGGCTGCGAGCGCGCGCTTGGCGACCGTGAAGTGGCC 849
Db 779 CTAGAGACCGCGCGCGCGAGTGGCTGCGAGCGCGCGCTTGGCGA-CCTGAGCACTGCC 836
Qy 850 GGATCCCGTGCACCGCTGGGACCGCAAGTGGC 881
Db 837 GGATCCATGCCCCCTGGACCCCAAGTTGGCCC 868

RESULT 7
LOCUS BI668879
DEFINITION 603294847F1 NIH_MGC_96 Homo sapiens CDNA clone IMAGE:5314354 5', mRNA linear EST 12-SEP-2001
ACCESSION BI668879
VERSION BI668879.1 GI:15583112
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 908)
NTH-MGC <http://mgi.ncl.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
CONTACT: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LUAM11796 row: m column: 11
High quality sequence stop: 782.
Location/Qualifiers
1. 908
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5314354"
/clone_lib="NIH_MGC_96"
/tissue_type="hypothalamus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-xhoI (gtcgag
); Oligo-dr primed using primer 5'-TTTTTTTTTTTTTTNN-3',
size-selected for average insert size 2.3 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 106 a 340 c 333 g 129 t
ORIGIN

Query Match 74.3%; Score 727.4; DB 13; Length 908;
Best Local Similarity 94.2%; Pred. No. 9.4e-112;
Matches 826; Conservative 5; Mismatches 36; Indels 10; Gaps 7;

Qy 22 AGTCGCGCGYMCRRGCGCGCTGCTGGGGCAGCATGGGGGTCCCGCTGCTCTGG 81
Db 6 AGTCGCGCGTCCGAGCGCGGCTCGCTGGGCGAGCATGGGGGTCCCGCTGCTCTGG 65
Qy 82 GGGCGCGCGGGCGGGCGCTGGCTTTTGGTGTGCTGTGCTGCTGCTGCTGCTGCTG 141
Db 66 GGGCGCGCGGGCGGGCGCTGGCTTTTGGTGTGCTGTGCTGCTGCTGCTGCTGCTG 125
Qy 142 CCCCCGCGCTCTGCGCGCGCGGTAAGAGAGCCCGCGGCTAAGCGCAGCGTCTCG 201
Db 126 CCCCCGCGCTCTGCGCGCGCGGTAAGAGAGCCCGCGGCTAAGCGCAGCGTCTCG 185
Qy 202 CCTTGTGCTGAGACTGCGCTCTCTGCGCTTCGCGGCTCAGTCCCGAGTGAGCG 261
Db 186 CCTTGTGCTGAGACTGCGCTCTCTGCGCTTCGCGGCTCAGTCCCGAGTGAGTG 245
Qy 262 CCGGGCGCGTGCAGGAGCTGCGCGCGCGCTGGCGCATCTGCTGGAGCGCGAGCTCAG 321
Db 246 CCGGGCGCGTGCAGGAGCTGCGCGCGCGCTGGCGCATCTGCTGGAGCGCGAGCTCAG 305
Qy 322 GAGCGCGCGCGCGCGAGGAGCTGAGGATCAGCAGCGCGCGCTCTTGGCGCAG 381
Db 306 GAG-GGGCGCGCGCGAGGAGCTGAGGATCAGCAGCGCGCGCTCTTGGCGCAG 364
Qy 382 CTGCTGCGCTGCTGGGGCGCGCGCGCAACTCTGATCCGGCTCTGGGCTTGGACGAC 441
Db 365 CTGCTGCGCTGCTGGGGCGCGCGCGCAACTCTGATCCGGCTCTGGGCTTGGACGAC 424
Qy 442 CCGGACGCGCTGCAGCGCAGCTGCTGCGCTGCTGCTGCGCGCGCGCTTGGACCTGCC 501
Db 425 CCGGACGCGCTGCAGCGCAGCTGCTGCGCTGCTGCTGCGCGCGCGCTTGGACCTGCC 484
Qy 502 GCGCTAGCAGCGCGCGCTTGTCCCGCGCGCGCT-CCCCCGCGCGCGCTTCCGACCGCGCC 560
Db 485 GCGCTAGCAGCGCGCGCTTGTCCCGCGCGCGCTTCCCGCGCGCGCTTCCGACCGCGCC 544
Qy 561 CCGGCTTACGACGAGCGCGCGCGCGCGCGCTGCTGAGGAGCGCGCGCGAGACACC 620
Db 545 CCGGCTTACGACGAGCGCGCGCGCGCGCGCTGCTGAGGAGCGCGCGCGAGACACC 604
Qy 621 CGAGCTGACCGCGCGCGCTGTTGAGTACTTCTGCGAGCGGATCTTGGGAGAGCGCGGA 680
Db 604 CGAGCTGACCGCGCGCGCTGTTGAGTACTTCTGCGAGCGGATCTTGGGAGAGCGCGGA 680

FEATURES
source
1. 908
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5314354"
/clone_lib="NIH_MGC_96"
/tissue_type="hypothalamus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-xhoI (gtcgag
); Oligo-dr primed using primer 5'-TTTTTTTTTTTTTTNN-3',
size-selected for average insert size 2.3 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."

QY 652 CTGGACGGAATTCTTGGCGGAAGCCGGAAGTCCGAGGGGGTGGCAGCCCCCGCCGCTC 711
|||||
Db 312 CTGGACGGAATTCTTGGCGGAAGCCGGAAGTCCGAGGGGGTGGCAGCCCCCGCCGCTC 253
|||||
QY 712 CCGCTGCGCCGACACGAGTGGCTCTGAGCTGCCCCCTGAGGGGGTGTGGGGGG 771
|||||
Db 252 CCGCTGCGCCGACACGAGTGGCTCTGAGCTGCCCCCTGAGGGGGTGTGGGGGG 193
|||||
QY 772 CTGCTGCGTGTAAACGCTAGAGACCCCGCGGCCCGAGTGCCTGACGCGCCCTCTTG 831
|||||
Db 192 CTGCTGCGTGTAAACGCTAGAGACCCCGCGGCCCGAGTGCCTGACGCGCCCTCTTG 133
|||||
QY 832 CCACCTGAGCACTCCCGGATCCCGTGCACCTGGGACCCAGAAAGTGCCTCCCGCATCC 891
|||||
Db 132 CCA-CTGAGCACTCCCGGATCCCGTGCAC-CTGGGACCCAGAAAGTGCCTCCCGCATCC 75
|||||
QY 892 CCGCACCAAGGACTGCTCCCGCCAGCAGCTCCAGAGCAACTTACCCCGCCAGCAGCC 951
|||||
Db 74 CCGCACCAAGGACTGCTCCCGCCAGCAGCTCCAGAGCAACTTACCCCGCCAGCAGCC 15
|||||
QY 952 TCTCACC 958
|||||
Db 14 TCTCACC 8
|||||

RESULT 12
BI668730
LOCUS
DEFINITION 603293168P1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5312320 5',
mRNA sequence.
ACCESSION BI668730
VERSION BI668730.1 GI:15582963
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 784)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: rcapbs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM11791 row: h column: 17
High quality sequence stop: 782.
Location/Qualifiers
1. .784
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5312320"
/clone.lib="NIH_MGC_96"
/tissue_type="hypothalamus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site:1: BamHI; Site:2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTNN-3',
size-selected for average insert size 2.3 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT 81 a 300 c 296 g 107 t
ORIGIN

Query Match 71.2%; Score 696.6; DB 13; Length 784;
Best Local Similarity 97.7%; Pred. No. 1.2e-106;
Matches 722; Conservative 5; Mismatches 9; Indels 3; Gaps 2;
QY 15 GCCAGCCAGTCCCGCGGCGGRRGCCCGCTCGCTGGGCGCAGATGGGGGGTGGCGCGT 74
|||||
Db 49 GCCAGCCAGTCCCGCGGCGGCGGCGGCTCGCTGGGCGCAGATGGGGGGTGGCGCGT 108
|||||
QY 75 GCTCTGGGGCGCGGGCGGGGGCGTGGGCTTTTGGTGTCTGCTGCTGGGCTGTT 134
|||||
Db 109 GCTCTGGGGCGCGGGCGGGGGCGTGGGCTTTTGGTGTCTGCTGCTGGGCTGTT 168
|||||
QY 135 TGGCGCGCGCGCGCGCGCTGTCGCGCGCGCGTAAAGAGAGCGCGCGCTTAAGCGCAGC 194
|||||
Db 169 TGGCGCGCGCGCGCGCGCTGTCGCGCGGGCGCGGTAAAGAGAGCGCGCGCTTAAGCGCAGC 228
|||||
QY 195 GTCTCCGCGCTTGGGTGAGACTGGCGCTCTCTCGCGCTTTCGGCGGTGAGTCCCGCAGG 254
|||||
Db 229 GTCTCCGCGCTTGGGTGAGACTGGCGCTCTCTCGCGCTTTCGGCGGTGAGTCCCGCAGG 288
|||||
QY 255 TGAGCGCGGGGGCGGTGCAGAGCTGGCGGGCGCTGGCGCATCTGCTGGAGGCCGA 314
|||||
Db 289 TGAGCGCGGGGGCGGTGCAGAGCTGGCGGGCGCTGGCGCATCTGCTGGAGGCCGA 348
|||||
QY 315 ACCTCAGGAGCGGGCGCGCGAGCGCGAGGCTGAGGATCAGCAGCGCGCTCT 374
|||||
Db 349 ACCTCAGGAGCGGGCGCGCGAGCGCGAGGCTGAGGATCAGCAGCGCGCTCT 408
|||||
QY 375 GCGCAGCTGTGCGCGCTGTCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 434
|||||
Db 409 GCGCAGCTGTGCGCGCTGTCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 468
|||||
QY 435 CGACGACCGCGCGCGCTGCGAGCGAGCTGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 494
|||||
Db 469 CGACGACCGCGCGCGCTGCGAGCGAGCTGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 528
|||||
QY 495 CCCTCGCGCGCTTAGCAGCGCGCGCTTGTCCCGCGCGCGCTTCCCGCGCGCGCGCGCGCGCG 554
|||||
Db 529 CCCTCGCGCGCTTAGCAGCGCGCGCTTGTCCCGCGCGCGCTTCCCGCGCGCGCGCGCGCGCG 588
|||||
QY 555 CCGCGCGCGCGCTTAGCAGCG 614
|||||
Db 589 CCGCGCGCGCGCTTAGCAGCG 648
|||||
QY 615 GACACCGCGCTGGACCGCGCGCTGCTGAGGTACTTGTGGGACGATTTCTTGGGGGAG 674
|||||
Db 649 GACACCGCGCTGGACCGCGCGCTGCTGAGGTACTTGTGGGACGATTTCTTGGGGGAG 708
|||||
QY 675 CGCGGACTCCGAGGGGGTGGCAGCG 734
|||||
Db 709 CGCGGACTCCGAGGGGGTGGCAGCG 765
|||||
QY 735 GGGCTCTGAGTGGCGCGCT 753
|||||
Db 766 GGGCTCTGAGTGGCGCGCT 784
|||||

RESULT 13
BI668730
LOCUS
DEFINITION BM715636 707 bp mRNA linear EST 28-FEB-2002
UI-E-EJ0-ahl-1-14-0-UI.r2 UI-E-EJ0 Homo sapiens cDNA clone
UI-E-EJ0-ahl-1-14-0-UI 5', mRNA sequence.
ACCESSION BM715636
VERSION BM715636.1 GI:19028894
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 707)
REFERENCE 1 (bases 1 to 707)
AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene

discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
The following repetitive elements were found in this cDNA
sequence: 256-326, >QC-rich#Low_complexity
Seq primer: M13 Reverse.

FEATURES
source
1..707
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-E-EJ0-ahi-1-14-0-UI"
/clone_lib="UI-E-EJ0"
/tissue_type="fetal eyes, lens, eye anterior segment,
optic nerve, retina, Retina Foveal and Macular, RPE and
Choroid"
/dev_stage="fetal and adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="organ: eye; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site_1: Ecor I; Site_2: Not I;
UI-E-EJ0 is a subtracted cDNA library constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an Ecor I adaptor, digested
with Not I, and cloned directionally into pT73-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for this library are: fetal eyes, AGAATCAAGA
; lens, CGATTACCGA; eye anterior segment, AATGCCGAT;
optic nerve, CCATTAAATG; retina, CCGCG; Retina Foveal and
Macular, GTC; RPE and Choroid, ACCTA. This library was
created for the program, Gene Discovery in the Visual
System, supported by National Eye Institute (NEI)."

BASE COUNT 91 a 285 c 240 g 90 t 1 others

Query Match 70.7%; Score 691.8; DB 14; Length 707;
Best Local Similarity 99.4%; Pred. No. 7.9e-106;
Matches 704; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 267 GCGGTCAGGAGCTGCGCGGCGCTGGCGATCTCTGGAGCGGACGTCAGGAGCG 326
|||||
Db 1 GCGGTCAGGAGCTGCGCGGCGCTGGCGATCTCTGGAGCGGACGTCAGGAGCG 60
|||||

QY 327 GCGCGGGCGGAGCGCAGGAGCTGAGGATCAGCAGCGCGCTCTGGCGGAGCTGCT 386
|||||
Db 61 GCGCGGGCGGAGCGCAGGAGCTGAGGATCAGCAGCGCGCTCTGGCGGAGCTGCT 120
|||||

QY 387 GCGGCTCTGGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 446
|||||
Db 121 GCGGCTCTGGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
|||||

QY 447 GCGGCTCTGAGCGACGCTGCTGCGGCTGCTGCGGCGGCGGCGGCGGCGGCGG 506
|||||
Db 181 GCGGCTCTGAGCGACGCTGCTGCGGCTGCTGCGGCGGCGGCGGCGGCGGCGG 240
|||||

QY 507 AGCAGCCAGCTGTGTCGCGCGCGCGCTCCCGCGCGCGCGCTCCGACCCCGCGCGGT 566
|||||
Db 241 AGCAGCCAGCTGTGTCGCGCGCGCGCTCCCGCGCGCGCGCTCCGACCCCGCGGT 300
|||||

QY 567 CTACGACGAGCGCGCGCGCGCGCGGATCTGAGGAGGAGCGAGACACACCCGACGT 626
|||||
Db 301 CTACGACGAGCGCGCGCGCGCGCGGATCTGAGGAGGAGCGAGACACACCCGACGT 360
|||||

QY 627 GGACCCCGAGCTGTTCAGGTACTTCTGGAGCGGATCTTGGGGAAGCGGAGCTCCGA 686
|||||
Db 361 GGACCCCGAGCTGTTCAGGTACTTCTGGAGCGGATCTTGGGGAAGCGGAGCTCCGA 420
|||||

QY 687 GGGGTGGCAGCGCGCGCGCGCGCTCCGCGGCTGCCCGACACGATGTGGGCTCTGAGCT 746
|||||
Db 421 GGGGTGGCAGCGCGCGCGCGCGCTCCGCGGCTGCCCGACACGATGTGGGCTCTGAGCT 480
|||||

QY 747 GCGCCCTGAGGGGCTGCTGGGGGCGCTGCTGTGTGAAACGCTAGAACCCCGCGGCC 806
|||||
Db 481 GCGCCCTGAGGGGCTGCTGGGGGCGCTGCTGTGTGAAACGCTAGAACCCCGCGGCC 540
|||||

QY 807 CAGGTGCTGACGCGCGCGCTCTTGGCACCTGACGACCTGCCGGATCCCGTGCACCCG 866
|||||
Db 541 CAGGTGCTGACGCGCGCGCTCTTGGCACCTGACGACCTGCCGGATCCCGTGCACCCG 600
|||||

QY 867 GGACCCAGAGTGCCTCCCGCGCATCCCGCACGAGGACTGCTCCCGCAGCAGCTCCAGA 926
|||||
Db 601 GGACCCAGAGTGCCTCCCGCGCATCCCGCACGAGGACTGCTCCCGCAGCAGCTCCAGA 660
|||||

QY 927 GCACTTACCCCGCGCGCGCGCGCTCTCACCGGAGATCCCTTACCCC 974
|||||
Db 661 GC-ACCTACCCCGCGCGCGCGCGCTCTCACCGGAGATCCCTTACCCC 707
|||||

RESULT 14
BI667137
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BI667137 771 bp mRNA linear EST 12-SEP-2001
603291530F1 NTH_MGC_96 Homo sapiens cDNA clone IMAGE:5310870 5',
mRNA sequence.
BI667137
BI667137.1 GI:15581370
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 771)
NIH-MGCS http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1787 row: 1 column: 07
High quality sequence stop: 763.

FEATURES
source
1..771
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5310870"
/clone_lib="NIH_MGC_96"
/tissue_type="hypothalamus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.3 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library

Result No.	Score	Query			DB	ID	Description
		Match	Length				
1	77.8	7.9	4403765	4	US-09-103-840A-2	Sequence 2, Appli	
C 2	71.6	7.3	4403765	4	US-09-103-840A-2	Sequence 2, Appli	
3	69.6	7.1	888	4	US-08-765-907A-2	Sequence 1, Appli	
C 4	69.6	7.1	2888	4	US-08-765-907A-1	Sequence 1, Appli	
5	69.4	7.1	4411529	4	US-09-103-840A-1	Sequence 4, Appli	
6	62.6	6.4	936	1	US-08-018-977C-4	Sequence 1, Appli	
C 7	60.4	6.2	4411529	4	US-09-103-840A-1	Sequence 1, Appli	
8	59.6	6.1	4488	4	US-08-400-030A-3	Sequence 3, Appli	
9	59.4	6.1	4524	2	US-08-843-998-7	Sequence 7, Appli	
10	59.4	6.1	4524	3	US-09-206-537-7	Sequence 7, Appli	
11	59.4	6.1	4524	4	US-09-430-854-7	Sequence 7, Appli	
12	59.2	6.0	4257	2	US-08-690-473-1	Sequence 1, Appli	
13	59.2	6.0	4257	4	US-09-259-821A-1	Sequence 1, Appli	
14	59.2	6.0	4257	4	US-08-843-659-1	Sequence 1, Appli	
C 15	59.2	6.0	12001	1	US-08-458-568A-11	Sequence 11, Appl	
16	58.8	6.0	8438	1	US-07-945-283-1	Sequence 1, Appli	
C 17	58.4	6.0	2823	1	US-08-398-008A-1	Sequence 1, Appli	
C 18	58.4	6.0	2823	2	US-08-893-333-1	Sequence 1, Appli	
C 19	58.2	5.9	30001	1	US-08-125-468-1	Sequence 1, Appli	
C 20	58.2	5.9	30001	2	US-08-474-933-1	Sequence 1, Appli	
21	58	5.9	44377	2	US-08-804-227C-7	Sequence 7, Appli	
22	58	5.9	44377	2	US-08-804-198-1	Sequence 1, Appli	
C 23	57.8	5.9	4257	2	US-08-690-473-1	Sequence 1, Appli	
C 24	57.8	5.9	4257	4	US-09-259-821A-1	Sequence 1, Appli	
C 25	57.8	5.9	4257	4	US-08-843-659-1	Sequence 1, Appli	
26	57.8	5.9	12001	1	US-08-458-568A-11	Sequence 11, Appl	
C 27	57.6	5.9	1030	3	US-08-858-003-2	Sequence 2, Appli	

; SEQ ID NO 2
; LENGTH: 888
; TYPE: DNA
; ORGANISM: Streptomyces pristinaespiralis
US-08-765-907A-2

Query Match 7.1%; Score 69.6; DB 4; Length 888;
Best Local Similarity 45.4%; Pred. No. 0.00049;
Matches 275; Conservative 3; Mismatches 327; Indels 1; Gaps 1;

QY 7 GGCACAGCCAGCCAGTCCGCGCGYMCRRGCCCGGCTCGTGGGCGCATGGCGGG 66
DB 92 GGGTGGCGGTGACCTGGCTGGAGTGGCCGGCGCGGTGCGGCGAGCGGGTGG 151
QY 67 TCGCCGCTCTCTGGGGGCGCGCGCGCGGCGTGGCGCTTTTGGTGTCTGTCTGCTC 126
DB 152 TGGCCGCTGATGCGCGCGCGCGCGCGAG - GCGGTGCGCGCGCTGCGCGCGGAC 210
QY 127 GGCCTGTTTGGCGCGCGCGCGCGCTCTCGCGCGCGCGGTAAGAGAGCCCGCGCCTA 186
DB 211 GTGGTGGTCTGCGCGTGGCGAGCGGTGGCGTGGAGCGGTGGAGTGTGGCGGG 270
QY 187 AGCGAGCGCTCTCGCGCTTGGCTGAGACTGGCGCTCTCGCGCTTCCGCGGTCACTG 246
DB 271 GTGATCGCGCGGTGCTGCGCGACACTTGTCTCGTCAAGAGCCGATCGCGGG 330
QY 247 CCCCAGGTGAGCGCGCGCGCGGTGCGAGAGCTGGCGCGCGCTGGCGCATCTGCTG 306
DB 331 CGGCTCGTGGAGCGCGCGCGGTGCGAGCGGTGGCGTGAACCGCATGTCGCGCCC 390
QY 307 GAGCGGAGCTCAGAGCGCGCGCGCGGTGCGAGCGGTGAGGATCAGCAGCGG 366
DB 391 TCGCTGGGTCTTACAGCGCGCGCGGTGCGCGCGGTGGTGTACCGACGCGCGGTG 450
QY 367 CGGCTCTCGCGCGCTGCTGCGCTGCTGGCGCGCGCGCGCAACTCTCATCGGCTCTG 426
DB 451 CGGCGCTGCTGAGCTGCTGGCGGTGGCGCGCGCGGTGGAGATGCGCGCGCGG 510
QY 427 GGTGGAGCAGCAGCCGCGCGCTGCGAGCAGCTGCTGCGCGCTCTGCTCCGCGCC 486
DB 511 CGGACAGCAGAGTACCGCGCGCGAGCGCGCGCGCGCGCGCGGTGCTGGCGCTTC 570
QY 487 CGGCTTACCTGCGCGCTAGAGCGCGCGGTGCTGCGCGCGCGCGCGCGCGCGCGG 546
DB 571 GGGCTGGGCTGGGTGAGCTGCTGCTGAGCTGGGCGCGCTGCGGACAGTGGCGCGG 630
QY 547 CTCCGACCGCGCGCGCTGACGACGCGCGCGCGCGCGCGATCTGAGGAGGCA 606
DB 631 CGGCATCTGGCGATGCTGGCGTCTGGCGCGCGCGCGCGCGCGAGGTGTAT 690
QY 607 GCGCAG 612
DB 691 TTCGAC 696

RESULT 4

US-08-765-907A-1/c
; Sequence 1, Application US/08765907A
; Patent No. 6352839
; GENERAL INFORMATION:
; APPLICANT: BLANC, Veronique
; APPLICANT: THIBAUT, Denis
; APPLICANT: BAMA-JACQUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSCHE, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: PARIS, Jean-Marc
; APPLICANT: DUTRUC-ROSSET, Gilles
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
; FILE REFERENCE: Mutasynthesis
; CURRENT FILING DATE: 1998-06-24

; CURRENT APPLICATION NUMBER: US/08/765,907A
; CURRENT FILING DATE: 1997-03-20
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2888
; TYPE: DNA
; ORGANISM: Streptomyces pristinaespiralis
US-08-765-907A-1

Query Match 7.1%; Score 69.6; DB 4; Length 2888;
Best Local Similarity 45.4%; Pred. No. 0.00047;
Matches 275; Conservative 3; Mismatches 327; Indels 1; Gaps 1;

QY 7 GGCACAGCCAGCCAGTCCGCGCGYMCRRGCCCGGCTCGTGGGCGCATGGCGGG 66
DB 1745 GGGTGGCGGTGACCTGGCTGGAGTGGCCGGCGCGGTGCGGCGAGCGGGTGG 1686
QY 67 TCGCCGCTCTCTGGGGGCGCGCGCGCGGTGCGCGCTTTTGGTGTCTGTCTGCTGCTC 126
DB 1685 TGGCCGCTGATGTCGCGCGCGCGCGCGAG - GCGGTGCGCGCGCTGCGCGCGGAC 1627
QY 127 GGCCTGTTTGGCGCGCGCGCGCGCTCTCGCGCGCGCGTAAAGAGAGCCCGCGCCTA 186
DB 1626 GTGGTGGTCTGCGGTGCGCGAGCGGTGGCGTGGAGGTGCTGGCGGG 1567
QY 187 AGCGAGCGCTCTCGCGCTTGGCTGAGACTGGCGCTCTCGCGCTTCCGCGCGGTCACTG 246
DB 1566 GTGATCGCGCGGTGCTGCGGTGCTGCGCGACACTTGTCTGCTCAAGAGCGGATCGCGGG 1507
QY 247 CCCCAGGTGAGCGCGCGCGCGGTGCGAGAGCTGGCGCGCGGTGCGCGCATCTGCTG 306
DB 1506 CGGCTCGTGGAGCGCGCGCGGTGCGAGCGGTGGGTGAACCGCATGTTTCGCCCC 1447
QY 307 GAGCGGAGCTCAGAGCGCGCGCGCGGTGCGCGCGCGGTGAGGATCAGCAGCGG 366
DB 1446 TCGCTGGGTCTTACAGGGCGCGCGGTGGTGTACCGAGCGCGCGGTG 1387
QY 367 CGGCTCTCGCGCGCTGCTGCGCGCTGCGCGCGCGCGCGCAACTCTGATCGCGCTCTG 426
DB 1386 CGGCGCTTGGTGGAGTGGTGGCGCGGTGGCGCGCGGTGGTGGAGATGCGCGCGGG 1327
QY 427 GGTGGAGCAGCAGCCCGCGCGCTGCGAGCAGCTGCTGCGCGCTCTGCTTCCGCGCC 486
DB 1326 CGGACAGCAGAGTACCGCGCGCGAGCGCGCGCGCATGCGCGCGGTGCTGGCGCTTC 1267
QY 487 CGGCTTACCTGCGCGCTAGCAGCGCGCTGCTGCGCGCGCGGTGCTGCGCGCGCGG 546
DB 1266 GGGCTGGGCTGGGTGAGCTGCTGGGTGGAGTGGGGGCGGTGGGAGACAGTGGCGCGG 1207
QY 547 CTCCGACCGCGCGCGCTTACGACGCGCGCGCGCGCGCGATGCTGAGGAGGCA 606
DB 1206 CGGCATCTGGCGATGCTGCGGTGCTGGCGCGCGCGATGCGCGCGGAGGTGTAT 1147
QY 607 GCGCAG 612
DB 1146 TTCGAC 1141

RESULT 5

US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-2007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 7.1%; Score 69.4; DB 4; Length 4411529;
Best Local Similarity 47.6%; Pred. No. 0.00037;
Matches 267; Conservative 0; Mismatches 291; Indels 3; Gaps 2;
Qy 258 GGGCGGGGGGGTGCAGAGAGTGGCGGGCGTGGCGATCTGTGGAGGCGCGAAGC 317
Db 337803 GGGCGGGGGGGGGCGGGCGGCTTGGCGTGGCGCGCGTGGCGCGTGGCGCGG 337862
Qy 318 TCAGGAGGGGGGGGGCGGCGGAGGCGAGGCTGAGGATCAGCAGGCGCGCTTGGC 377
Db 337863 GTTGGCGAGAGCGGGCGGCGCGCGCGCCACCGGTTCGCGGGCGCGCGAAGGA 337922
Qy 378 GCAGCTGCTGCGCGCTCTGGGGCGCGCGCGCGCAACTCTGATCGGCTCTGGGCTTGGACGA 437
Db 337923 CCCGCCGCG-GCGCGCGCGACCGCGCGCGCGCGCGATGAGCAGACCGCGCTTTCGCGCG 337981
Qy 438 CGACCGCGAGCGCGCTGCAGCGCAGTGGCTGCGGCTGTCTCGCGCGCGCGCGCTTGAACC 497
Db 337982 CGCGCGCGCGCGCGCGCGCGTTCGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 338041
Qy 498 TGCGCGCGCTAGCAGCGCGAGCTTGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 557
Db 338042 CGAACAGTGCCTGCGCGCTGCGCTCGATTCGATCCCGCGCGCGCGCGCGCGCGCGCG 338101
Qy 558 GCGCGCGCTTACAGCAGCG 617
Db 338102 ATCGCGCGCTGCG 338161
Qy 618 ACCGCGAGTGCAGCG 677
Db 338162 CG 338221
Qy 678 GGACTTCGAGGGGTGGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 737
Db 338222 CGCGCGCGAGACCG 338281
Qy 738 CTCTGAGTGCCTTGCAGGCGCGTGTGGGGCGCGTGTCTGCGTGTGAACCGCTTAGAGAC 797
Db 338282 ATCGCGCGGGCG 338339
Qy 798 CCCGGCG 818
Db 338340 GCG 338360

RESULT 6

US-08-018-977C-4
; Sequence 4, Application US/08018977C
; Patent No. 5686601
; GENERAL INFORMATION:
; APPLICANT: Weber, Peter C.
; TITLE OF INVENTION: DNA Sequences Encoding Mutant Antiviral
; TITLE OF INVENTION: Regulatory Proteins
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Pennsylvania State University
; STREET: 113 Technology Center
; CITY: University Park
; STATE: PA
; COUNTRY: US
; ZIP: 16802-7000
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 720 Kb storage
; COMPUTER: IBM PS/2

; OPERATING SYSTEM: MS-DOS 6.22
; SOFTWARE: Microsoft Word for Windows 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/018,977C
; FILING DATE: 18-FEB-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/726,071
; FILING DATE: 05-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Thomas J. Monahan
; REGISTRATION NUMBER: 29,835
; REFERENCE/DOCKET NUMBER: 91-1039/2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (814) 865-6277
; TELEFAX: (814) 865-3591
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 936 bases
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
US-08-018-977C-4

Query Match 6.4%; Score 62.6; DB 1; Length 936;
Best Local Similarity 44.0%; Pred. No. 0.0071;
Matches 399; Conservative 4; Mismatches 488; Indels 16; Gaps 3;
Qy 25 CCGCGCGYMCRRRCCTCGGCTGGGGGAGCATGGGGGGTGGCGCGCTGCTGTGGCGGCC 84
Db 20 CCGCGCGCGAGCAGCG 79
Qy 85 CCGCGCGCGCGCGCGCTGGGCTTTTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 144
Db 80 GCGGGGGGTCTCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 139
Qy 145 CCGCGCGTCTGCGCGCGCGCGGTAAAGGAGCGCCCGCGCGCTTAAGCGCAGCTCTCCGCC 204
Db 140 CCATCGCGCGCACGCGGACAGTAACACCAACACCAACACCAACACCAACAGCAGCGCG 199
Qy 205 TTGGCTGAGACTGGCGCTCCTCGCGCGCTTCGGCGGTGAGTGGCGCGCGCGCGCGCG 264
Db 200 GCTCCCGCGAGTCTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 259
Qy 265 GGGCGGTGCGAGGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 324
Db 260 GGGTTGGGTTGGGTTGTTGAACGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 314
Qy 325 GGGCGCGCGCGGAGCG 384
Db 315 CCGCTTGTCAACAGACCG 374
Qy 385 CTGCGCGCTGGGGCG 440
Db 375 CTCGCCCGCGCGCTCTCCCGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 434
Qy 441 CCGCGAGCGCGCTGCAGCGCGAGCTCGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCG 500
Db 435 CCGCGCGCGCGCGCGCTCG 494
Qy 501 CGCGCTAGCAGCGCGCTTGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 560
Db 495 CCGCTGCGTGGAGCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 554
Qy 561 CCGCGTCTAGCAGCG 620
Db 555 GCG 614
Qy 621 CGAGCTGGAGCG 680
Db 615 GCG 674


```
QY 681 CTCGAGGGGGTGCAGCCCGCGCGCTCCGCGTGGCGCGGACCATGATGTGGCTC 740
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 675 GGGGCGGGGGTGGAGGGTGGGACGGGCGCTCCCGCGGCGCGCCCTCCGCGGCCCG 734
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 741 TGAGCTGCCCTGAGGGGGTGTGGGGGGCTGTGCTGCTGTGAACGCTAGAGACCC 800
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 735 CCGGCTCCCTCCGCGCGCTCTCTCAGCAGGAGGGCGGGTCCGTCGAGGAGA- 790
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 801 GGGGCCCCAGGTGCTGCAGCGCGCTCTTGCCACCTGAGCACTGCCGGATGCCGTG 860
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 791 ---GGCGGGGTGGGCCAGGAAACCTCCCGCAGTCCAGCGCTCCCGCCCTCGCG 847
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 861 ACCCTGGAGCCAGAGTGCCTCCCGCATCCGCGCACAGGACTGCTCCCGCCACAG 920
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 848 CGCGAGGGCCAGAGGGCGGACGACACCCCTCCGACTCAGGGCGCGGGGGCGCG 907
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 921 TCCAGAG 927
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 908 GCCAGG 914
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6284328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 6.2%; Score 60.4; DB 4; Length 4411529;
Best Local Similarity 46.4%; Pred. No. 0.011;
Matches 365; Conservative 0; Mismatches 408; Indels 14; Gaps 4;

QY 191 CAGCGTCTCGGCGCTGGCTGAGACTGGCGCTCTCGCGCTTCCGCGGTGAGTCCCC 250
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3936062 CGGCGCGCGCTCGCTGCCACGCGCTGTGCTTCCGCTTCCGCGGTGAGTCCCC 3936003

QY 251 GAGGTGAGCGGGGGGGGCTGCAGAGCTGGCGGGGCGTGGCGCATCTGCTGGAGG 310
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3936002 CGGTCCCGCGGTGCGCGCGCGCGCTGTGATGCCCGCGCGCGGTGCGCGGACCC 3935943

QY 311 CCGAAGCTCAGAGCGGGCGGGCGCGAGAGCTGAGGATCAGCAGCGCGCG 370
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3935942 CGCCTTGGCGCGCTTGGCGCGCGGGAAGCGCTTGCCTTGGGAGAGGGCGCGTGT 3935883

QY 371 TCTTGGCGAGCTGTGCGCGCTTGGGGGCGCGCGCACTGATCGCGCTGCTGGGT 430
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3935882 CGCGGCGCGCGCGCTGC-----CGCGCGCGCGCGCGCGCGATACCGCGGTGCC 3935827

QY 431 TGAGCAGCAGCCGCGCGCTGCAGCGAGCTGCTCGCGCTGTGCTCGCGCGCGCC 490
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3935826 CGGTTCCGCGCGGTGTGGCCAGCGCGCGAGCGCGCGGTGCGCGCGCTCCGCGGTG 3935767

QY 491 TTGACCTGCGCGCTTACGAGCGCCAGCTTGTCCCCCGCGCGCTCCCGCGCGCTCC 550
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3935766 CGCGCGGACCGCGCGCAAGAGCGCTGCGCTGCGCGCGGTGCGCGCGCGCGCGCC 3935707

QY 551 GACCGCGCGCGCGCTGTACGAGCAGCGCGCGCGCGCGCGCGATGCTGAGGAGCGCG 610
```

```
Db 3935706 CCGCGCTTCCCGCGCTACCGCTGGCGCGCTGGCTAGCCCGGTGCTGTTCAGGCGCGCT 3935647
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 611 ACAGAGACACCCGACGAGTCCCGGAGCTGTGAGGTACTTGTGGGACGAGTTCCTTCCGG 670
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3935646 TTG-----CCGCTTGGCGCGGTGGCGCGGACCACTTGGCGCGCTCAATGCTG 3935593
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 671 GAAGCCGCGACTCCGAGGGGTGGAGCCCGCGCGCGCTTCGCCCTGCGCGCGGACGACG 730
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3935592 GGGTCCCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3935533
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 731 ATGTGGGCTCTGA-GCTGCGCGCTGAGGGGCTGCTGGGGCGCTGCTGCGTGTGAACGC 789
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3935532 CGGTCCCGGGAAACACCGCGCGGATCCGGGTTTGGCGCGCGCGCGCGCGCGCT 3935473
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 790 CTAGAGACCCCGCGCGCGCGCGCTGCTGCACGCGCGCTTGTGCACCTGAGCACTGCCCC 849
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3935472 TGGCGCGCTCACCCTTGGCGCGCGCGCGCGCGCTGCTGCCACGCGCTGCTTCCG 3935413
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 850 GGATCCCGTGCACCTGGGACCGAGAGTGCCCGCGCGCGCTTCCGCCACCGAGGACTGCTCC 909
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3935412 TTCTGGCGCTCACCACCGCGCGCGCGCGCTGCGCGGTGCGCGCGCGCGCGCGCG 3935356
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 910 CGCCGAGCAGTCCAGAGCAACTTACCCCGCGCGCGCGCGCGCTTCCCGCGAGATCCCT 969
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3935355 CGCGCGCGCGCGCGCTTCCCGCGCGCGCGCGCGCTTGGCGCGCGCTTGGCGCGCGCGCGCGCGCG 3935296
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 970 ACCCGCT 976
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3935295 TTGCGT 3935289
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
US-08-406-030A-3/c
; Sequence 3, Application US/08406030A
; Patent No. 6270989
; GENERAL INFORMATION:
; APPLICANT: Treco, Douglas A.
; APPLICANT: Heartlein, Michael W.
; APPLICANT: Hauge, Brian M.
; APPLICANT: Selden, Richard F.
; TITLE OF INVENTION: Protein Production and Delivery
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.c.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/406,030A
; FILING DATE: 17-MAR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/243,391
; FILING DATE: 13-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,586
; FILING DATE: 03-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/911,533
; FILING DATE: 10-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/787,840
; FILING DATE: 05-NOV-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/789,188
```


Db 708 CCGGGGGGGGCTCTGGGGGGCTCCGGCGACCTTACCCCGCATATGACAGCCTGGGC 767
QY 461 AGCTCGCTCGCTGCTCCGGCGCGCTTGACCTTCGGCCCTAGCAGCCACCTTG 520
Db 768 CACCTGTGCGACCCCGGGGGCGCGCATGAACATGCCGTCCGGCTGCCGACCCC 827
QY 521 TCCCGCGCGCTCCCGCGGGGGCGCTCGGACCCCGGGCGCGGCTTACGAGAGCGGC 580
Db 828 GGGCTGTGGCGGGCGCGCGACCGCGGCGACGCGCGCGCGCGCGCGCGCGCG 887
QY 581 CCGC-GGGCGCGATCTGAGAGGCGAGCGAGACACCGACCTGACCCCGAGCTG 639
Db 888 GGGCAGTGGCAGCGCATCGGGCGGGCGCGTGGTGGCGCAGCGGGCTGGCGTCC 947
QY 640 TTGAGTACTTGTGGGCGGATTTCTGGGGAAGCGGACTTCCGAGGGGTGGCAGCC 699
Db 948 ATCTGCGACTCGACACCGCGCGAGCTCGAGGCGTTCGCGAGCGCTTCAAGCAG 1007
QY 700 CCGCGCGCGCTCGCGCGTCCGGCGACCATCCCGCCAGGACTGCTCCCGCGCAGCAGCTTCA 934
Db 1008 CCGCGCATCAAGTGGCGGTGACGACCGCGCGCTGAGCTTTCACCGCGCGCGAGAGCGC 1067
QY 755 AGGGCGTGTGGGGCGCTGCTGCTGTAAGCGCTAGAGACCCCGCGCGCGCGCG 814
Db 1068 ATCCCGGGGTGGGTCTACTAGCCAGACCATCTCGAGGCTTCAGTCTGCTCAGCTC 1127
QY 815 CCGCGCGCGCTTCTGCGCACCTTCTGCGGAGCGCGCTGCGCGCGCGCGCGCG 874
Db 1128 TCGCACACACATGATCGCGCTCAAGCCCATCTCGAGGCTGGCTCGAGAGGCGCGAG 1187
QY 875 AGTGTCCCGCGCATCCCGCCAGGACTGCTCCCGCGCAGCAGCTCCAGAGCAACTTA 934
Db 1188 GCGCGCCAGCGGAGAAATGAACAGCTGAGCTTTCAACGGCGCGCGAGAGAGCGC 1247
QY 935 CCGCGCGCGCGCGCGCTTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 979
Db 1248 AGCGGACTTCCATCGC-CGCGCGCGAGAGCGCTCCCTCGAGGC 1291

RESULT 10

US-09-206-537-7
Sequence 7, Application US/09206537
Patent No. 6130052

GENERAL INFORMATION:

APPLICANT: Van Baren, Nicolas
APPLICANT: Coulie, Pierre G.
APPLICANT: De Smet, Charles
APPLICANT: Lucas, Sophie
APPLICANT: Boon, Thierry

TITLE OF INVENTION: LEUKEMIA ASSOCIATED GENES

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:

ADDRESSEE: Wolf, Greenfield & Sacks, P.C.

STREET: 600 Atlantic Avenue

CITY: Boston

STATE: MA

COUNTRY: US

ZIP: 02210

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/206,537

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/845,998

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Van Amsterdam, John R.

REGISTRATION NUMBER: 40,212
REFERENCE/DOCKET NUMBER: L0461/7008

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)720-3500

TELEFAX: (617)720-2441

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 4524 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

FEATURE:

NAME/KEY: CDS

LOCATION: 174..1433

US-09-206-537-7

Query Match 6.1%; Score 59.4; DB 3; Length 4524;

Best Local Similarity 46.0%; Pred. No. 0.022;

Matches 352; Conservative 0; Mismatches 401; Indels 12; Gaps 4;

QY 221 CTCCTCGCGCTTCCGGCGGTGCTGCTGAGGCGGAGGTCAGCGGGGGGGGGTTCAGGAGC 280

Db 533 CTCCTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 592

QY 281 TGGCGGGCGGTGGCGGCTGCTGCTGAGGCGGAACTCAGGAGCGGCGCGCGCGCGGAGG 340

Db 593 GGGCGGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 652

QY 341 CGCAGGAGCTGAGGATCAGCAGCGCGCTCTGTCGCGAGCTGCTGCTGCTGCTGCTGCTG 400

Db 653 CGCGCGCGCGCGCGGAGGCGGCTGGCG-----GCCCGGGGGCGCGCGCGCGCGCG 707

QY 401 CCCCCCGCACTTGATCGGGCTTGGGCTTGGACACGACCCCGAGCGGCGCTTCAGCGC 460

Db 708 CGGGCGGGGGCTCTTGGCGGCTCCCGCGGCTCCCGCGACCTTCACCCGCTATGACAGCCTGG 767

QY 461 AGCTCGCTCGCGCTGCTTCCCGCGCGCTTACCCCTGCGCGCTTACCGAGCGCGCGCGCTT 520

Db 768 CACCTGTGCGACCCCGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 827

QY 521 TCCCGCGCGCTCCCGCGGGCGCTCCGACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 580

Db 828 GGGCTGTGGCGGGCGGCGGCGACACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 887

QY 581 CCGC-GGGCGCGGATGCTGAGGAGCGGAGCGGAGCAGACACCGACGTCGACCCCGAGCTG 639

Db 888 GGGCAGTGGCAGCGCATCGCGCGCGCGCGCGGCGGCTGGTGGGCGCGAGCGCGCGCTGG 947

QY 640 TTGAGGTACTTGTGGGAGCGGATTTCTTGGGGAAGCGCGGAGCTCCGAGGGGTGGCGAGC 699

Db 948 ATCTGCGACTCGGACACGAGCGGCGGCGGAGCTCGAGGCGCTTCCGAGAGCGCTTCAAGCAG 1007

QY 700 CCGCGCGCGCTCCCGCGTCCCGCGACCATGATGTTGGGCTCTGAGCTG-----CCCCCTG 754

Db 1008 CCGCGCATCAAGCTGGCGGTGACGACCGCGGCGGCTGGGCTCGGCGCTGGCCAACTCAAG 1067

QY 755 AGGGCGTGTGGGGCGCTGCTGCTGTAAGCGCTAGAGACCCCGCGCGCGCGCGCGCGGTCG 814

Db 1068 ATCCCGGGGTGGGTCTACTAGCCAGACCATCTGAGGTTTCAGTCTGCTCAGCTC 1127

QY 815 CCGCGCGCGCTTCTTGGCACCTTCTGCGGAGCGGCTGCGGAGCTCCCGTTCGACCTTGGAGCC 874

Db 1128 TCGCACACACATGATCGCGCTCAAGCCCATCTCGAGGCTGGCTCGAGAGGCGCGAG 1187

QY 875 AAGTGTCCCGCGCATCCCGCCAGGACTGCTCCCGCGCAGCAGCTCCAGAGCAACTTA 934

Db 1188 GCGCGCCAGCGGAGAAATGAACAGCTGAGCTTTCAACGGCGCGCGAGAGAGCGC 1247

QY 935 CCGCGCGCGCGCGCGCTTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 979

Db 1248 AAGCGGACTTCCATCGC-CGCGCCGAGAGGGCTCCCTCGAGGC 1291

RESULT 11

US-09-430-854-7

; Sequence 7, Application US/09430854

; Patent No. 6271019

; GENERAL INFORMATION:

; APPLICANT: Van Baren, Nicolas

; APPLICANT: Coulie, Pierre G.

; APPLICANT: De Smet, Charles

; APPLICANT: Lucas, Sophie

; APPLICANT: Boon, Thierry

; TITLE OF INVENTION: LEUKEMIA ASSOCIATED GENES

; NUMBER OF SEQUENCES: 16

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.

; STREET: 600 Atlantic Avenue

; CITY: Boston

; STATE: MA

; ZIP: 02210

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/430,854

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/845,998

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Van Amsterdam, John R.

; REGISTRATION NUMBER: 40,212

; REFERENCE/DOCKET NUMBER: L0461/7008

; TELEPHONE: (617)720-3500

; TELEFAX: (617)720-2441

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 4524 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; MOLECULE TYPE: linear

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 174..1433

; US-09-430-854-7

Query Match 6.1%; Score 59.4; DB 4; Length 4524;

Best Local Similarity 46.0%; Pred. No. 0.022;

Matches 352; Conservative 0; Mismatches 401; Indels 12; Gaps 4;

QY 221 CTCCTCCGCGCTTCCGCGGTTCAGTCCCGAGGTGAGCGCGCGCGGTGCGAGGC 280

Db 533 CTCCTCCGCGCTTCCGCGGTTCAGTCCCGAGGTGAGCGCGCGCGGTGCGAGGC 592

QY 281 TGGCGCGCGCGCTTCCGCGGTTCAGTCCCGAGGTGAGCGCGCGGTGCGAGGC 340

Db 593 CGGCGCGCGCGCTTCCGCGGTTCAGTCCCGAGGTGAGCGCGCGGTGCGAGGC 652

QY 341 CGCAGGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGC 400

Db 653 CGGCGCGCGCGCTTCCGCGGTTCAGTCCCGAGGTGAGCGCGCGGTGCGAGGC 707

QY 401 CCCCCCGCACTCTGATCCCGCTCTGCGGTTCAGTCCCGAGGTGAGCGCGGTGCGAGGC 460

Db 708 CGGCGCGCGCGCTTCCGCGGTTCAGTCCCGAGGTGAGCGCGCGGTGCGAGGC 767

QY 461 AGCTCGCTCGGCGCTTCTGCTCGCGCGCGCGCTTCTGACCTTGGCGCCCTAGCAGCCAGCTTG 520

Db 768 CACCTGTGCGACCCCGCGCGCGCGCGCGCTTCTGACCTTGGCGCCCTAGCAGCCAGCTTG 827

QY 521 TCCCGCGCGCGCTTCCGCGGTTCAGTCCCGAGGTGAGCGCGCGGTGCGAGGC 580

Db 828 GGGTGTGCGCGCGCGCGCGCGCTTCTGACCTTGGCGCCCTAGCAGCCAGCTTG 887

QY 581 CGGC-GGGCGCGGTGCTGAGGAGCGCGCGCGGTGAGCGCGGTGCGAGGC 639

Db 888 GGGAGGTGCGCGCGCGCGCGCTTCTGACCTTGGCGCCCTAGCAGCCAGCTTG 947

QY 640 TTGAGGTACTTCTGCGCGCGCGCGCTTCTGACCTTGGCGCCCTAGCAGCCAGCTTG 699

Db 948 ATCTGCGACTCGGACCGGCGCGCGCTTCTGACCTTGGCGCCCTAGCAGCCAGCTTG 1007

QY 700 CGGCGCGCGCTTCCGCGGTTCAGTCCCGAGGTGAGCGCGGTGCGAGGC 754

Db 1008 CGGCGCATCAAGCTGGCGGTGAGCGCGCGGTGAGCGCGGTGCGAGGC 1067

QY 755 AGGCGGTGCTGGCGCGCTTCTGACCTTGGCGCCCTAGCAGCCAGCTTG 814

Db 1068 ATCCCGCGCGCTTCTGCGCGCGCGCTTCTGACCTTGGCGCCCTAGCAGCCAGCTTG 1127

QY 815 CTGACCGCGCGCTTCTGCGCGCGCGCTTCTGACCTTGGCGCCCTAGCAGCCAGCTTG 874

Db 1128 TCGCACAACAATGATGCGCGCTTCTGACCTTGGCGCCCTAGCAGCCAGCTTG 1187

QY 875 AAGTGCCTCCCGCGGTTCAGTCCCGAGGTGAGCGCGGTGCGAGGC 934

Db 1188 GCGCGCGCGCGGTGAGCGCGGTGAGCGCGGTGAGCGCGGTGAGCGGC 1247

QY 935 CCGCGCGCGCGCTTCTGACCTTGGCGCCCTAGCAGCCAGCTTG 979

Db 1248 AAGCGGACTTCCATCGC-CGCGCCGAGAGGGCTCCCTCGAGGC 1291

RESULT 12

US-08-690-473-1

; Sequence 1, Application US/08690473

; Patent No. 5878923

; GENERAL INFORMATION:

; APPLICANT: Leopardi, Rosario

; APPLICANT: Roizman, Bernard

; TITLE OF INVENTION: HERPES SIMPLEX VIRUS ICP4 AS AN

; TITLE OF INVENTION: INHIBITOR OF APOPTOSIS

; NUMBER OF SEQUENCES: 2

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Arnold, White & Durkee

; STREET: P.O. Box 4433

; CITY: Houston

; STATE: Texas

; COUNTRY: USA

; ZIP: 77210

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/690,473

; FILING DATE: 26-JUL-1996

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Highlander, Steven L.

; REGISTRATION NUMBER: 37,642

; REFERENCE/DOCKET NUMBER: ARCD:239

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 512/418-3000

; TELEFAX: 512/474-7577

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 4257 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-690-473-1

Query Match 6.0%; Score 59.2; DB 2; Length 4257;
Best Local Similarity 44.5%; Pred. No. 0.024;
Matches 374; Conservative 0; Mismatches 453; Indels 13; Gaps 3;

```
QY 48 CTGGGCGAGCATGCGGGTCCGCGTCTGCGGGCGCGGGCGGGGGCGGCGCTCGGCGCT 107
Db 2307 CGGCGGCGAGGAGCCCGTGGCGCGCGTGGCGCGCGTGGCGCGCGTGGCGCGCGCT 2366
QY 108 TTGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 167
Db 2367 GGGCGCGCGCTGCGGGAGCGCGCTGCGGAGCTCGCGGCGCGCGCGCGCGCGGGA 2426
QY 168 AAGGAGCGCGCGCGCTAAGCGAGCGTCTCGCGCGCTGCGCGCGCTGCGCGCGCTCG 227
Db 2427 CCGTCTTTTACAAACAGAGCGCTGCGCGCGCTGCGCGCGCTGCGCGCGCTGCGCG 2486
QY 228 CCGCTTCCGCGGTCAGTCCCGAGTGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 287
Db 2487 CGCGCGCGAGCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2542
QY 288 GCGCTGCGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 347
Db 2543 AGCGCAAGAGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2602
QY 348 GCGTGGAGTACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 407
Db 2603 CGAAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 467
QY 408 CAACTCTGATCCGCGTCTGCGCTTGGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCG 2720
Db 2663 CGCGCGCGTCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2840
QY 588 CCGCGATGCTGAGGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 647
Db 2841 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2900
QY 648 CTTGCTGGGAGGATCTTGGGAGCGCGGAGTCCGAGGGGTGCGAGCGCGCGCGCGCG 707
Db 2901 CCGCGTCTCCCGTCCCGTGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2953
QY 708 CTTCCGCGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 767
Db 2954 TGGCGCTGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3013
QY 768 GCGCGTCTGCTGTAAGCGCGCTAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 827
Db 3014 GCGCGAGGAGGAGGATTAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3073
QY 828 CTTGCCACCGTACGACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 887
Db 3074 CCGTGGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3133
```

RESULT 13

US-09-259-821A-1
; Sequence 1, Application us/09259821A
; Patent No. 6210926
; GENERAL INFORMATION:

; APPLICANT: LEOPARDI, ROSARIO
; APPLICANT: ROIZMAN, BERNARD
; TITLE OF INVENTION: HERPES SIMPLEX VIRUS ICP4 IS AN INHIBITOR OF APOPTOSIS
; FILE REFERENCE: ARCD:317
; CURRENT APPLICATION NUMBER: US/09/259,821A
; CURRENT FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: 08/690,473
; PRIOR FILING DATE: 1996-07-26
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4257
; TYPE: DNA
; ORGANISM: HERPES VIRUS, TYPE 1
US-09-259-821A-1

Query Match 6.0%; Score 59.2; DB 4; Length 4257;
Best Local Similarity 44.5%; Pred. No. 0.024;
Matches 374; Conservative 0; Mismatches 453; Indels 13; Gaps 3;

```
QY 48 CTGGGCGAGCATGCGGGTCCGCGTCTGCGGGCGCGGGCGGGGGCGGCGCTCGGCGCT 107
Db 2307 CGGCGGCGAGGAGCCCGTGGCGCGCGTGGCGCGCGTGGCGCGCGTGGCGCGCGCT 2366
QY 108 TTGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 167
Db 2367 GGGCGCGCGCTGCGGGAGCGCGCTGCGGAGCTCGCGGCGCGCGCGCGCGCGGGA 2426
QY 168 AAGGAGCGCGCGCGCTAAGCGAGCGTCTCGCGCGCTGCGCGCGCTGCGCGCGCTCG 227
Db 2427 CCGTCTTTTACAAACAGAGCGCTGCGCGCGCTGCGCGCGCTGCGCGCGCTGCGCG 2486
QY 228 CCGCTTCCGCGGTCAGTCCCGAGTGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 287
Db 2487 CGCGCGCGAGCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2542
QY 288 GCGCTGCGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 347
Db 2543 AGCGCAAGAGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2602
QY 348 GCGTGGAGTACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 407
Db 2603 CGAAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 467
QY 408 CAACTCTGATCCGCGTCTGCGCTTGGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCG 2720
Db 2663 CGCGCGCGTCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2840
QY 468 TCGCGCTCTGCTGCGCGCGCGCGCTGACCTGCGCGCGCTAGCAGCGCGCGCGCGCG 527
Db 2721 GCGCGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2780
QY 528 GCGCGTCCCGCGCGCGCGTCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 587
Db 2781 CCGCGACCGCTGGCGGGCTGGCGGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 2840
QY 588 CCGCGATGCTGAGGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 647
Db 2841 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2900
QY 648 CTTGCTGGGAGGATCTTGGGAGCGCGGAGTCCGAGGGGTGCGAGCGCGCGCGCGCG 707
Db 2901 CCGCGTCTCCCGTCCCGTGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2953
QY 708 CTTCCGCGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 767
Db 2954 TGGCGCTGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3013
QY 768 GCGCGTCTGCTGTAAGCGCGCTAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 827
Db 3014 GCGCGAGGAGGAGGATTAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3073
QY 828 CTTGCCACCGTACGACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 887
```

Db 3074 CCTTGGCGCCCTCGGCGCCGCTGCGCCGCGATGCGCGCCTGGATGCGCCAGATCCCGGACC 3133

RESULT 14
US-08-843-659-1
; Sequence 1, Application US/08843659
; Patent No. 6218103
; GENERAL INFORMATION:
; APPLICANT: Leopardi, Roasrio
; APPLICANT: Roizman, Bernard
; TITLE OF INVENTION: HERPES SIMPLEX VIRUS US3 AND ICP4 AS
; TITLE OF INVENTION: INHIBITORS OF APOPTOSIS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESS: Arnold, White & Durkee
; CITY: Houston
; STATE: Texas
; COUNTRY: United States
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Highlander, Steven L.
; REGISTRATION NUMBER: 37,642
; REFERENCE/DOCKET NUMBER: ARSB:519
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4257 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-843-659-1

Query Match 6.0%; Score 59.2; DB 4; Length 4257;
Best Local Similarity 44.5%; Pred. NO. 0.024;
Matches 374; Conservative 0; Mismatches 453; Indels 13; Gaps 3;

QY 48 CTGGGGGACGATGGGGGGTGGCGGCTGCTCTGGGGGCGCGGGGCGGGGGGGGGGGGGGGGGCT 107
Db 2307 CGGCGGCGAGCGAGCGCGCGCGTGGCGCGCGTGGCGCGCGTGGCGCGCGTGGCGCGCGT 2366

QY 108 TTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 167
Db 2367 GGGCGCGCGCTGCGCGGGGACCGCGCGCTGCGGAGCTGCGGGCGCGCGCGCGCGCGGA 2426

QY 168 AAGAGAGCCCGCGCGCTTAAGCGCAGCGCTCTCCGCGCTTGGCTGAGACTGGCGCTCCCTCG 227
Db 2427 CTTGCTGTTTACAAACAGAGCTGCGCCCTCTGCGCGCGGGCGCGCGCGCGCGCGGA 2486

QY 228 CCGCTTCGGCGGTCAGTGGCCCGAGGTGAGCGCGCGGGGGGGGGTGGCAGGAGCTGGCGCG 287
Db 2487 GCGCGCGCGAGCGCTGCGCGCGCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2542

QY 288 GCGCTGCGCGCTGCTGCTGAGGCGCAAGTTCAGGAGCGCGCGCGCGCGCGCGCGCGCGGA 347
Db 2543 AGCGCAAGAGTCCCGGA 2602

QY 348 GCGTGGAGTACAGAGCGCGCGCTCTTGGCGCAGCTGCTGCGCGCTGCTGGGCGCGCGCGCG 407
Db 2603 CGAAGAGAGCGCGCGCGCGCGCGCTCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2662

QY 408 CAATCTGTATCGCGGCTCTGGGCTTGGACGACGACACCGCGCGCTTGCAGCGAGCTCGC 467
Db 2663 CGCCCGCTTCCAGCGCGCGCGCG--GGCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 2720

QY 468 TCGCGCTCTGCTCGCGCGCGCGCTTGCACCTTGCAGCGCTAGCAGCCAGCTTGTCTCCCG 527
Db 2721 GCG 2780

QY 528 GCGCGTCCCGCGCGCGCGCTCCGACCGCGCGCGCGCGCGCGCGCTACGACGAGCGCGCGG 587
Db 2781 CCGCGACCGCTTGGCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2840

QY 588 CCGCGATGCTGAGGAGCGAGCGACGACACCGCGCGCGCGCGCGCGCGCTTGTAGGTA 647
Db 2841 CGCGCGCGCGCGCTTGGAGCGCTACTGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCG 2900

QY 648 CTTGCTGGGACGATCTTGGCGGAGCGCGGACTCCGAGGCGGCTGCGAGCGCGCGCGCG 707
Db 2901 CCGCGTGTTCG 2953

QY 708 CCTCGCGCTGCGCGCGCGCGCGCGCGCGCTGAGCTGCGCGCGCGCGCGCGCGCGCGCG 767
Db 2954 TGGCGCTGATCG 3013

QY 768 GCGCGTGTGCTGTGAACGCTAGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 827
Db 3014 GCGCGGAGGAGCAGCATAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3073

QY 828 CTTGCGACCGCTGAGCAGCTGCGCGGATCCGCTGACCGCTGGGACCGAGAGTGCCTCCCG 887
Db 3074 CCCTGCGCGCTGCG 3133

RESULT 15
US-08-458-568A-11/c
; Sequence 11, Application US/08458568A
; Patent No. 5821339
; GENERAL INFORMATION:
; APPLICANT: Schaffer, Priscilla A.
; APPLICANT: Yeh, Lily
; TITLE OF INVENTION: Compositions and Methods for Treatment of Herpesvirus
; TITLE OF INVENTION: Infections
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5821339rls
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,568A
; FILING DATE: 02-JUNE-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/065,146
; FILING DATE: 05-MAY-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Leary Ph.D., Kathryn R.
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: DFCI-0029
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12001 base pairs

Job time : 9767 secs

```

; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Herpes simplex virus
; STRAIN: Herpes Simplex Virus Type 1
; US-08-458-568A-11

Query Match      6.0%; Score 59.2; DB 1; Length 12001;
Best Local Similarity 44.5%; Pred. No. 0.023;
Matches 374; Conservative 0; Mismatches 453; Indels 13; Gaps 3;

QY 48 CTGGGGCAGCATGGCGGGTGCOCGTGCTCTGGGGGCGCCCGGGGGCCCTCGGCCT 107
Db 4367 CGCGCGCAGCAGCGCCCGTGGCGCGCGTGGCGCGTGGTGGCGGGGCCCT 4308

QY 108 TTTGGTGTCTGTCTGCTCGGCCTGTTTCGGCGCGCCCGCCCGCGCTGCGCGCGCGCGGT 167
Db 4307 GGGCGCCCGCGCTGCCCGGGACCGCGCCCTGCCGAGTCCGCGCGCCCGCCCGCGCGGA 4248

QY 168 AAGGAGCCCCCGCGCTAAGCGCAGCGTCTCGGCCCTTGGCTGAGACTGGCGCTCCTCG 227
Db 4247 CCTGCTGTTGACACACAGACCTTGCGCCCTCTGTGGCGCGCGCGCCAGCGACCGGA 4188

QY 228 CCGCTTCCGGCGTCACTGCTCCCGAGGTGAGCGCGCGGGCGGTGCAGGAGCTGGCGCG 287
Db 4187 CGCGCGCAGCGCTGCGCGCGCGCGCGCTCCGCGCGCGCGCGGAGG---GGCGCA 4132

QY 288 GCGCTGGCGCATCTGTGAGGCGCGACGTGAGGAGCGGGCGGGCGCGAGGCGGAGGA 347
Db 4131 AGCGCAAGAGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGA 4072

QY 348 GGCTGAGGATCAGCAGCGCGCGCTCTGCGCGAGCTGCTGCGCGCTCTGGGGCGCCCCCG 407
Db 4071 CGAAGAGAGCGCGCGGAGCGCCCGCGGTGCGACGCGCGCGCGCGCGCGCGCGCG 4012

QY 408 CAACTCTGATCCGGCTCTGGGCTTGGACGACACCGCGCGCGCTGACGCGAGCTCGC 467
Db 4011 CGCGCCCGCTCCAGCGCGCGCG---GGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCG 3954

QY 468 TCGGCTCTGCTCGGCGCGCGCTTGACCTGCGCGCGCTGAGCGCGCGAGTTGTCCCGCG 527
Db 3953 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3894

QY 528 GCGCGTCCCGCGCGCGCGCTCGGACCGCGCGCGCGCGCGCTGACGAGCGCGCGCGCG 587
Db 3893 CCGCGACCCCTGGCGCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3834

QY 588 CCGGATGCTGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 647
Db 3833 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3774

QY 648 CTTGCTGGGAGGATTTTCGGGAGCGCGCGGACTCGGAGGGGTGGCGCGCGCGCGCG 707
Db 3773 CCGCGCTGTTCCCGCTCCCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3721

QY 708 CTTCCGCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 767
Db 3720 TGGCTGATGCGCGCGCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3661

QY 768 GCGCGTCTGCTGTAAGCGCGCTAGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 827
Db 3660 GCGCGCAGCAGCAGATACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3601

QY 828 CTTGCCACCTGAGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 887
Db 3600 CCCTGGCGCGCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3541
```


GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 12, 2003, 18:14:02 ; Search time 151 Seconds
(without alignments)
5687.068 Million cell updates/set

Title: US-09-803-589-5
Perfect score: 979
Sequence: 1 gaattcgacagggccagc.....gaggatccctaccctccggc 979

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 593429 seqs, 438583890 residues

Total number of hits satisfying chosen parameters: 1186858

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA.*

- 1: /cgn2.6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2.6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2.6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2.6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2.6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2.6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2.6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2.6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2.6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 10: /cgn2.6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
- 11: /cgn2.6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 12: /cgn2.6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
- 13: /cgn2.6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 14: /cgn2.6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %		Length	DB	ID	Description
1	977	99.8	979	10	US-09-803-589-5		Sequence 5, Appli
2	912.8	93.2	1050	9	US-09-984-271-58		Sequence 58, Appl
3	507.8	51.9	1027	10	US-09-803-589-13		Sequence 13, Appl
c	4	276.6	28.3	301	10	US-09-860-107-1701	Sequence 1701, Appl
5	274.4	28.0	349	10	US-09-960-352-71		Sequence 71, Appl
6	244.8	25.0	372	10	US-09-983-965-2663		Sequence 2663, Appl
7	167	17.1	342	10	US-09-983-965-3684		Sequence 3684, Appl
8	75.2	7.7	1614	9	US-09-976-740-45		Sequence 45, Appl
9	75.2	7.7	1614	12	US-10-023-529-45		Sequence 45, Appl
10	75.2	7.7	1614	12	US-10-023-523-45		Sequence 45, Appl
11	75.2	7.7	12425	9	US-09-976-740-50		Sequence 50, Appl
12	75.2	7.7	12425	12	US-10-023-529-50		Sequence 50, Appl
13	75.2	7.7	12425	12	US-10-023-523-50		Sequence 50, Appl
14	73.4	7.5	2561	9	US-09-976-740-48		Sequence 48, Appl
15	73.4	7.5	2561	12	US-10-023-529-48		Sequence 48, Appl
16	73.4	7.5	2561	12	US-10-023-523-48		Sequence 48, Appl
17	69.4	7.1	2307	9	US-09-893-519A-87		Sequence 87, Appl
c	18	63.2	6.5	12425	9	US-09-976-740-50	Sequence 50, Appl
c	19	63.2	6.5	12425	12	US-10-023-529-50	Sequence 50, Appl

c	20	63.2	6.5	12425	12	US-10-023-523-50	Sequence 50, Appl
c	21	62.8	6.4	12733	9	US-10-032-393-47	Sequence 47, Appl
c	22	62.8	6.4	12739	9	US-10-032-393-8	Sequence 8, Appl1
c	23	62.4	6.4	4653	9	US-10-101-388-2	Sequence 2, Appl1
c	24	62.4	6.4	8036	9	US-10-101-388-1	Sequence 1, Appl1
c	25	62	6.3	790	9	US-10-004-717-20	Sequence 20, Appl
c	26	60.4	6.2	1614	9	US-09-976-740-45	Sequence 45, Appl
c	27	60.4	6.2	1614	12	US-10-023-529-45	Sequence 45, Appl
c	28	60.4	6.2	1614	12	US-10-023-523-45	Sequence 45, Appl
c	29	59.2	6.0	4257	9	US-09-825-288A-1	Sequence 1, Appl1
c	30	58.2	5.9	2529	10	US-09-815-242-7852	Sequence 7852, Ap
c	31	58.2	5.9	37113	9	US-09-999-121-11	Sequence 11, Appl
c	32	58.2	5.9	43058	10	US-09-954-456-292	Sequence 292, App
c	33	58.2	5.9	43058	10	US-09-954-456-529	Sequence 529, App
c	34	58.2	5.9	43058	10	US-09-880-107-3950	Sequence 3950, Ap
c	35	57.8	5.9	2561	9	US-09-976-740-48	Sequence 48, Appl
c	36	57.8	5.9	2561	12	US-10-023-529-48	Sequence 48, Appl
c	37	57.8	5.9	2561	12	US-10-023-523-48	Sequence 48, Appl
c	38	57.8	5.9	4257	9	US-09-825-288A-1	Sequence 1, Appl1
c	39	57.6	5.9	1030	9	US-09-735-056-2	Sequence 2, Appl1
c	40	57.6	5.9	41936	10	US-09-967-768A-116	Sequence 116, App
c	41	57.4	5.9	88421	9	US-09-976-059-1	Sequence 1, Appl1
c	42	57	5.8	1758	10	US-09-815-242-7825	Sequence 7825, Ap
c	43	57	5.8	1817	10	US-09-924-417-64	Sequence 64, Appl
c	44	56.4	5.8	987	10	US-09-772-656-5	Sequence 5, Appl1
c	45	56.4	5.8	987	10	US-09-772-656-9	Sequence 9, Appl1

ALIGNMENTS

RESULT 1
US-09-803-589-5
; Sequence 5, Application US/09803589
; Patent No. US20020112251A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Goodearl, Andrew D.J.
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC AND OTHER
; TITLE OF INVENTION: USES
; FILE REFERENCE: 07334-325001
; CURRENT APPLICATION NUMBER: US/09/803,589
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: US 09/128,709
; PRIOR FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: US 60/054,645
; PRIOR FILING DATE: 1997-08-04
; PRIOR APPLICATION NUMBER: US 09/130,491
; PRIOR FILING DATE: 1998-08-06
; PRIOR APPLICATION NUMBER: US 60/054,966
; PRIOR FILING DATE: 1997-08-06
; PRIOR APPLICATION NUMBER: US 60/058,108
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: US 09/388,280
; PRIOR FILING DATE: 1999-09-01
; PRIOR APPLICATION NUMBER: US 09/388,279
; PRIOR FILING DATE: 1999-09-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 979
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (58)...(837)
US-09-803-589-5

Query Match 99.8%; Score 977; DB 10; Length 979;
Best Local Similarity 100.0%; Pred. No. 1.2e-208;
Matches 979; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 GAATTCGCACGAGCCAGCCAGTCCTCCGCCSGTMCORRRCCCGGCTCGCTGGGGCAGCATG 60
|
Db 1 GAATTCGCACGAGCCAGCCAGTCCTCCGCCSGTMCORRRCCCGGCTCGCTGGGGCAGCATG 60
|
QY 61 GCGGGTCCGCGCTCTCTGGGGCCGCGGCGCGGGGCGTGGCTTTTGGTGTCTGCTG 120
|
Db 61 GCGGGTCCGCGCTCTCTGGGGCCGCGGCGCGGGGCGTGGCTTTTGGTGTCTGCTG 120
|
QY 121 CTGTCGCGCCCTTTTGGCCGCCGCCCGCTCTGCGCGCGCGCGGTAAGAGAGCCCGC 180
|
Db 121 CTGTCGCGCCCTTTTGGCCGCCGCCCGCTCTGCGCGCGCGCGGTAAGAGAGCCCGC 180
|
QY 191 GGCCTAAGCGAGCGTCTCCGCCCTTGCTGAGACTGCGCTCTCTCGCCGCTTCGGGGG 240
|
Db 191 GGCCTAAGCGAGCGTCTCCGCCCTTGCTGAGACTGCGCTCTCTCGCCGCTTCGGGGG 240
|
QY 241 TCAGTGCCTCCGAGGTAGCGCGCGGGGCGTGCAGAGCTGGCGGCGCTGGCGCAT 300
|
Db 241 TCAGTGCCTCCGAGGTAGCGCGCGGGGCGTGCAGAGCTGGCGGCGCTGGCGCAT 300
|
QY 301 CTGTCGAGGCGCGAAGCTCAGAGCGGCGCGCGCGCGAGCGAGGCTGAGGATCAG 360
|
Db 301 CTGTCGAGGCGCGAAGCTCAGAGCGGCGCGCGCGAGCGAGGCTGAGGATCAG 360
|
QY 361 CAGGCGCGCTCTGGCGCAGCTGTGGCGCTCTGGGGCGCGCGCGCGCAACTCTGATCCG 420
|
Db 361 CAGGCGCGCTCTGGCGCAGCTGTGGCGCTCTGGGGCGCGCGCGCGCAACTCTGATCCG 420
|
QY 421 GCTCTGGCTTGGACGAGACCGCGCGCTGCAGCGCAGCTTGTCCCGCGCGCGTCTGCTC 480
|
Db 421 GCTCTGGCTTGGACGAGACCGCGCGCTGCAGCGCAGCTTGTCTGGCTCTGCTC 480
|
QY 481 CGCGCGCGCTTGCCTGCGCGCTAGCAGCGCGAGCTTGTCCCGCGCGCGTCCCGCGC 540
|
Db 481 CGCGCGCGCTTGCCTGCGCGCTAGCAGCGCGAGCTTGTCCCGCGCGCGTCCCGCGC 540
|
QY 541 CGCGCGCTCCGACCGCGCGCTTACGACGCGCGCGCGCGCGCGCGCGGATGCTGAG 600
|
Db 541 CGCGCGCTCCGACCGCGCGCTTACGACGCGCGCGCGCGCGCGCGCGGATGCTGAG 600
|
QY 601 GAGCAGCGCAGACACCGCGCTGACCGCGAGCTGTTGAGGTACTTCTGGGAGCG 660
|
Db 601 GAGCAGCGCAGACACCGCGCTGACCGCGAGCTGTTGAGGTACTTCTGGGAGCG 660
|
QY 661 ATTCTTGGGGAAGCGGGAATCCGAGGGGTGGCAGCGCGCGCGCTCCGCGCTGCC 720
|
Db 661 ATTCTTGGGGAAGCGGGAATCCGAGGGGTGGCAGCGCGCGCGCTCCGCGCTGCC 720
|
QY 721 GCCGACACGATGTGGGCTGTGAGCTGCCCTGAGGGCTGTGGGGCGCTGCTGCGT 780
|
Db 721 GCCGACACGATGTGGGCTGTGAGCTGCCCTGAGGGCTGTGGGGCGCTGCTGCGT 780
|
QY 781 GTGAAGCGCTTAGAGACCGCGCGCCAGTGCTGCAGCGCGCTTGTGCCACCTGA 840
|
Db 781 GTGAAGCGCTTAGAGACCGCGCGCCAGTGCTGCAGCGCGCTTGTGCCACCTGA 840
|
QY 841 GCATGCCCGGATCCCGTGCACCTGGGAGCCAGAAAGTGCCTCCCGCGCATCCCGCACAG 900
|
Db 841 GCATGCCCGGATCCCGTGCACCTGGGAGCCAGAAAGTGCCTCCCGCGCATCCCGCACAG 900
|
QY 901 GACTGTCCCGCGCAGACGTTCAGAGAACTTACCCGCGCGCAGCGCTCTCACCG 960
|
Db 901 GACTGTCCCGCGCAGACGTTCAGAGAACTTACCCGCGCGCAGCGCTCTCACCG 960
|
QY 961 AGGATCCCTACCCCTGGC 979
|
Db 961 AGGATCCCTACCCCTGGC 979
|
```

RESULT 2
US-09-984-271-58
; Sequence 58, Application US/09984271

```
; Publication No. US20030040088A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 71 Human Secreted Proteins
; FILE REFERENCE: P2030P1
; CURRENT APPLICATION NUMBER: US/09/984,271
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 09/482,273
; PRIOR FILING DATE: 2000-01-13
; PRIOR APPLICATION NUMBER: PCT/US99/15849
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: 60/092,921
; PRIOR FILING DATE: 1998-07-15
; PRIOR APPLICATION NUMBER: 60/092,922
; PRIOR FILING DATE: 1998-07-15
; PRIOR APPLICATION NUMBER: 60/092,956
; PRIOR FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 267
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 58
; LENGTH: 1050
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-984-271-58
```

```
Query Match 93.2%; Score 912.8; DB 9; Length 1050;
Best Local Similarity 98.9%; Pred. No. 2e-194;
Matches 955; Conservative 5; Mismatches 2; Indels 4; Gaps 4;

QY 14 GGCAGCAGTCGCCCGSGYMCRRGCGCGCTCGCTGGGGCAGCATGCGGGGTGCGCGC 73
|
Db 11 GGCAGCAGTCGCCCGCGCTCGCGAGCGCGGCTCGCTGGGGCAGCATGCGGGGTGCGCGC 70
|
QY 74 TGCCTCTGGGGCCGCGGCGCGGCGCTGCGCGCGCGGTAAGAGAGCCCGCGGCTTAAGCGAG 193
|
Db 71 TGCCTCTGGGGCCGCGGCGCGGCGCTGCGCGCGCGGTAAGAGAGCCCGCGGCTTAAGCGAG 190
|
QY 134 TTCGGCGCGCGCGCGCTGTCGCGCGCGCGGTAAGAGAGCCCGCGGCTTAAGCGAG 193
|
Db 131 TTCGGCGCGCGCGCGCTGTCGCGCGCGCGGTAAGAGAGCCCGCGGCTTAAGCGAG 190
|
QY 194 CGTCTCGCGCTTGGCTGAGACTGGCGCTCTCGCGCTTCCGGGGTCACTGCGCGCGC 253
|
Db 191 CGTCTCGCGCTTGGCT-AGACTGGCGCTCTCGCGCTTCCGGGGTCACTGCGCGCGC 249
|
QY 254 GTGAGCGCGCGCGCGCTGCGAGGCTGGCGCGCGCTGCGCGCTGCTGCTGAGAGCGC 313
|
Db 250 GTGAGCGCGCGCGCGCTGCA-GACTGGCGCGCGCTGCGCGCTGCTGCTGGA-GCCG 307
|
QY 314 AACCTCAGGAGCGCGCGCGCGAGCGCGCGCGCGAGGCTGAGGATCAGCAGCGCGCGTCC 373
|
Db 308 AACCTCAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTCC 366
|
QY 374 TGGCGCAGCTGCTGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCGCGCTGG 433
|
Db 367 TGGCGCAGCTGCTGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCGCGCTGG 426
|
QY 434 ACAGCAGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 493
|
Db 427 ACAGCAGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 486
|
QY 494 ACCCTGCGCGCTAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCGCG 553
|
Db 487 ACCCTGCGCGCTAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCGCG 546
|
QY 554 CCGCGCGCGCGCTACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 613
|
Db 547 CCGCGCGCGCGCTACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 606
|
QY 614 AGACACCGAGCTGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 673
|
Db 607 AGACACCGAGCTGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 666
|
```


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

; TITLE

; TITLE

```

RESULT 9
US-10-023-529-45
; Sequence 45, Application US/10023529
; Patent No. US020129388A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY
; TITLE OF INVENTION: PROTEINS AND T

```


QY 790 CTAGAGACCCCGCGCCAGGTGCC 815
Db 676 CGCCCGCGCGCAGCCACAGCGCC 701

RESULT 11

US-09-976-740-50
; Sequence 50, Application US/09976740
; Publication No. US20020194633A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/976,740
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 50
; LENGTH: 12425
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-976-740-50

Query Match 7.7%; Score 75.2; DB 9; Length 12425;
Best Local Similarity 48.6%; Pred. No. 9.5e-09;
Matches 275; Conservative 0; Mismatches 278; Indels 13; Gaps 2;
QY 261 GCGCGGGGGTGCAGAGCTGGCGGCGTGGCGCATCTGCTGAGGCGCGAAGCTCA 320
Db 2969 GCGCGGGCGGACCTGGAGCGCATCTGCGGATGGTGGCGGCGCAGCGCCGAGGC 3028
QY 321 GGAGCGGGCGCGCGCGAGCGCAGGAGCTGAGGATCAGAGCGCGCGCTCTG----- 375
Db 3029 GGAGCGCAGCGCGCGCGAGCTCGAGAACTGATCCAGCAGCGCGCTGCTCGGGTTCAG 3088
QY 376 -----GCGCAGCTGCTGCGGCTCTGGGGCGCCCGCGCAACTCTGATCGGCTCTGGGC 429
Db 3089 CTACAAGGGAGCATCTCGTACCGCAACGCGCGCGCTCCAGCGCCCGCGCGGAGC 3148
QY 430 TTGAGCAGACACCCCGACGCGCTGCAGGCGAGCTCGCTCGCGCTCTGCTCCGCGCCGC 489
Db 3149 CACCCGCGGGCG 3208
QY 490 CTTGACCTTCCGCGCTAGAGCGCCAGCTTGTCCCGCGCGCGCTCCCGCGCGCGCGCTC 549
Db 3209 GCGCGCGCGCACCG 3266
QY 550 CGACCCCGCGCGCGCGCTTACAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 609
Db 3267 GCCCGGGCG 3326
QY 610 GAGGAGACACCCCGACCTGGACCGCGAGCTTGTGAGGTACTTGTGGAGCGGATCTTGGC 669
Db 3327 CAGCCGGCG 3386
QY 670 GGAAGCGCGAGCTCCAGAGGGGTGGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 729
Db 3387 GCTCCCGCGCGTGGCG 3446
QY 730 GATGTGGGCTCTAGCTGCCCCCTGAGAGGGGTGCTGGGGGCGCTGCTGTGGTGTGAACGC 789

Db 3447 GCCCGGAGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3506
QY 790 CTAGAGACCCCGCGCCAGGTGCC 815
Db 3507 CGCCCGCGCGCAGCCACAGCGCC 3532

RESULT 12

US-10-023-529-50
; Sequence 50, Application US/10023529
; Patent No. US20020129388A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,529
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 50
; LENGTH: 12425
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-023-529-50

Query Match 7.7%; Score 75.2; DB 12; Length 12425;
Best Local Similarity 48.6%; Pred. No. 9.5e-09;
Matches 275; Conservative 0; Mismatches 278; Indels 13; Gaps 2;
QY 261 GCGCGGGGGTGCAGAGCTGGCGGCGTGGCGCATCTGCTGAGGCGCGAAGCTCA 320
Db 2969 GCGCGGGCGGACCTGGAGCGCATCTGCGGATGGTGGCGGCGCAGCGCCGAGGC 3028
QY 321 GGAGCGGGCGCGCGCGAGCGCAGGAGCTGAGGATCAGAGCGCGCGCTCTG----- 375
Db 3029 GGAGCGCAGCGCGCGCGAGCTCGAGAACTGATCCAGCAGCGCGCTGCTCGGGTTCAG 3088
QY 376 -----GCGCAGCTGCTGCGGCTCTGGGGCGCCCGCGCAACTCTGATCGGCTCTGGGC 429
Db 3089 CTACAAGGGAGCATCTCGTACCGCAACGCGCGCGCTCCAGCGCCCGCGCGGAGC 3148
QY 430 TTGAGCAGACACCCCGACGCGCTGCAGGCGAGCTCGCTCGCGCTCTGCTCCGCGCCGC 489
Db 3149 CACCCGCGGGCG 3208
QY 490 CTTGACCTTCCGCGCTAGAGCGCCAGCTTGTCCCGCGCGCGCTCCCGCGCGCGCGCTC 549
Db 3209 GCGCGCGCGCACCG 3266
QY 550 CGACCCCGCGCGCGCGCTTACAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 609
Db 3267 GCCCGGGCG 3326
QY 610 GAGGAGACACCCCGACCTGGACCGCGAGCTTGTGAGGTACTTGTGGAGCGGATCTTGGC 669
Db 3327 CAGCCGGCG 3386

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 12, 2003, 19:29:20 ; Search time 1589 Seconds
(without alignments)
9978.214 Million cell updates/sec

Title: US-09-803-589-5
Perfect score: 979
Sequence: 1 gaattcgacagggcagc.....gaggtacctaccctcgtgc 979

Scoring table: OLIGO_NUC

Capop 60.0 , Capext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 15

Total number of hits satisfying chosen parameters: 95542

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- EST:*
- 1: em_estba:*
 - 2: em_esthum:*
 - 3: em_estin:*
 - 4: em_estnu:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_esttc:*
 - 9: gb_est1:*
 - 10: gb_est2:*
 - 11: gb_hic:*
 - 12: gb_est3:*
 - 13: gb_est4:*
 - 14: gb_est5:*
 - 15: em_estfun:*
 - 16: em_estcom:*
 - 17: gb_gss:*
 - 18: em_gss_hum:*
 - 19: em_gss_inv:*
 - 20: em_gss_pin:*
 - 21: em_gss_vrt:*
 - 22: em_gss_fun:*
 - 23: em_gss_mam:*
 - 24: em_gss_mus:*
 - 25: em_gss_other:*
 - 26: em_gss_pro:*
 - 27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	612	62.5	707	14	BM715636
2	561	57.3	784	13	BI668730
3	517	52.8	839	13	BI546579
4	514	52.5	764	13	BI669134
5	504	51.5	570	12	BF438085
6	487	49.7	589	14	BM876317

7	484	49.4	599	14	BM695925
8	483	49.3	537	14	BM720023
c 9	478	48.8	593	13	BM504127
c 10	474	48.4	601	13	BI711080
c 11	472	48.2	515	13	BM052768
c 12	472	48.2	518	13	BI789216
c 13	472	48.2	541	12	BF724971
c 14	472	48.2	562	13	BM142527
c 15	472	48.2	569	13	BI789206
c 16	472	48.2	573	13	BI789119
c 17	472	48.2	589	13	BM129392
c 18	472	48.2	590	13	BI789105
c 19	472	48.2	596	13	BM129323
c 20	471	48.1	713	14	BM674078
c 21	467	47.7	619	14	BQ268412
c 22	464	47.4	476	10	BE045542
c 23	464	47.4	498	14	BQ581370
c 24	463	47.3	692	13	BI490760
c 25	462	47.2	631	12	BG108317
c 26	459	46.9	586	13	BM129592
c 27	454	46.4	626	13	BM508023
c 28	453	46.3	504	12	BF725537
c 29	453	46.3	581	13	BI963705
c 30	451	46.1	559	13	BM504353
c 31	450	46.0	817	13	BI668984
c 32	449	45.9	604	13	BM508441
c 33	448	45.8	617	13	BI711823
c 34	444	45.4	469	9	AI796185
c 35	444	45.4	469	9	BQ102329
c 36	443	45.3	469	9	AI261620
c 37	442	45.1	600	13	BM129679
c 38	439	44.8	584	14	BQ362055
c 39	434	44.3	475	13	BM510525
c 40	434	44.3	482	13	BM511355
c 41	433	44.2	771	13	BI667137
c 42	431	44.0	481	12	BG819901
c 43	431	44.0	517	13	BM510394
c 44	431	44.0	878	13	BI603988
c 45	430	43.9	690	12	BG716263

ALIGNMENTS

RESULT 1	BM715636	707 bp	mRNA	linear	EST 28-FEB-2002
LOCUS	UI-E-EJ0-ahi-l-14-0-UI-r2	UI-E-EJ0	Homo sapiens	cdna	clone
DEFINITION	UI-E-EJ0-ahi-l-14-0-UI 5', mRNA sequence.				
ACCESSION	BM715636				
VERSION	BM715636.1	GI:19028894			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 707)				
TITLE	Bonaldo,M.F., Lennon,G. and Soares,M.B.				
JOURNAL	Normalization and subtraction: two approaches to facilitate gene discovery				
MEDLINE	Genome Res. 6 (9), 791-806 (1996)				
COMMENT	97044477				
	Contact: Soares, MB				
	Program for Rat Gene Discovery and Mapping				
	University of Iowa				
	451 Eckstein Medical Research Building Iowa City, IA 52242, USA				
	Tel: 319 335 8230				
	Fax: 319 335 9565				
	Email: msoares@blue.weeg.uiowa.edu				
	Tissue Procurement: Dr. Gregg Hageman				
	cdna Library preparation: Dr. M. Bento Soares, University of Iowa				
	cdna Library Arrayed by: Dr. M. Bento Soares, University of Iowa				
	DNA Sequencing by: Dr. M. Bento Soares, University of Iowa				


```
Db 132 CGGTGCGCCCTTTGGTGTCTGCTGCTCGGCTGTGCTGCGCCGCCCGCCGCTCTGGC 191
QY 158 CGCGCGCGGTAAAGAGAGCCCGCGGCTTAAGCGCAGCGTCTCGCCCTTGGCTGAGACTG 217
Db 192 CGCGCGCGGTAAAGAGAGCCCGCGGCTTAAGCGCAGCGTCTCGCCCTTGGCTGAGACTG 251
QY 218 GCCTCTCGCCCTTCCGCGGCTAGTCCCGAGGTGAGCGCGCGGGCGGCTGCAGG 277
Db 252 GCCTCTCGCCCTTCCGCGGCTAGTCCCGAGGTGAGCGCGGGCGGCTGCAGG 311
QY 278 AGCTGCGCGCGGCTGGCGCATCTGCTGAGAGCGCAAGCTCAGGAGCGCGCGCGCGCG 337
Db 312 AGCTGCGCGCGGCTGGCGCATCTGCTGAGAGCGCAAGCTCAGGAGCGCGCGCGCGCG 371
QY 338 AGCGCAGGAGGTGAGGATCAGCAGCGCGGCTCTGCGCGCAGCTGCTGCGGCTTGGG 397
Db 372 AGCGCAGGAGGTGAGGATCAGCAGCGCGGCTCTGCGCGCAGCTGCTGCGGCTTGGG 431
QY 398 CGCGCCCGCCCACTCTGATCCGGCTTGGGCTTGGAGCAGCAGCCCGAGCGCCCTGCAG 457
Db 432 CGCGCCCGCCCACTCTGATCCGGCTTGGGCTTGGAGCAGCAGCCCGAGCGCCCTGCAG 491
QY 458 CGCAGCTCGCTCGGCTCTGCTCGCGCGGCTTGGAGCGCGGCTTGGAGCGCGCGCGAGC 517
Db 492 CGCAGCTCGCTCGGCTCTGCTCGCGCGGCTTGGAGCGCGGCTTGGAGCGCGCGAGC 551
QY 518 TTGTCCTCCGCGCGGCTCCCGCGCGGCTTGGAGCGCGGCTTGGAGCGCGCGAGC 577
Db 552 TTGTCCTCCGCGCGGCTCCCGCGCGGCTTGGAGCGCGGCTTGGAGCGCGCGAGC 611
QY 578 GCGCGCGCGCGCGGATGCTGAGAGCGAGCGCGCGAGCAGCAGCCCGAGTGGAGCCCGAGC 637
Db 612 GCGCGCGCGCGCGGATGCTGAGAGCGAGCGCGAGCAGCAGCAGCAGCAGCAGCAGCAGC 671
QY 638 TGTGAGGTACTTGTGAGGAGGATCTTGGGGAAGCGCGGACTCCGAGGGGGTGGCAG 697
Db 672 TGTGAGGTACTTGTGAGGAGGATCTTGGGGAAGCGCGGACTCCGAGGGGGTGGCAG 731
QY 698 CCC 700
Db 732 CCC 734

RESULT 3
BI546579
LOCUS
DEFINITION
603191602F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5262672 5',
mRNA sequence.
ACCESSION BI546579
VERSION BI546579.1 GI:15433891
KEYWORDS
SOURCE EST.
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 839)
NIH-MGC http://img.cni.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1162 row: d column: 01
High quality sequence stop: 793.
Location/Qualifiers
```

```
source 1. 839
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5262672"
/clone_lib="NIH_MGC_95"
/tissue_type="hippocampus"
/lab_host="DH10b"
/note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.5 kb and
normalized to 10^5. This is a primary library enriched
for full-length clones and constructed using the
cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 87 a 311 c 324 g 117 t
ORIGIN

Query Match 52.8%; Score 517; DB 13; Length 839;
Best Local Similarity 99.7%; Pred. No. 2.2e-226;
Matches 617; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 38 GCCCGGCTCGCTGGGCGAGCATGGGGGTCGCGGCTGCTCTGGGGCCCGGGCGGGG 97
Db 22 GCCCGGCTCGCTGGGCGAGCATGGGGGTCGCGGCTGCTCTGGGGCCCGGGCGGGG 81
QY 98 GCCTCGCGCTTTTGGGTGCTGCTGCTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 157
Db 82 GCCTCGCGCTTTTGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 141
QY 158 CGCGCGCGGTAAAGAGAGCCCGCGGCTTAAGCGCAGCGTCTCGCGCCCTTGGCTGAGACTG 217
Db 142 CGCGCGCGGTAAAGAGAGCCCGCGGCTTAAGCGCAGCGTCTCGCGCCCTTGGCTGAGACTG 201
QY 218 GCCTCTCGCCCTTCCGCGGCTCAGTGCCTCGAGGTGAGGCGGGGGCGGTGCGAG 277
Db 202 GCCTCTCGCCCTTCCGCGGCTCAGTGCCTCGAGGTGAGGCGGGGGCGGTGCGAG 261
QY 278 AGCTGCGCGCGGCTGGCGCATCTGCTGAGGCGCGAAGCTCAGGAGCGGGCGGGCGG 337
Db 262 AGCTGCGCGGCGCTGGCGCATCTGCTGAGGCGCGAAGCTCAGGAGCGGGCGGGCGG 321
QY 338 AGCGCAGGAGGTGAGGATCAGCAGCGCGGCTCTGCGCGCAGCTGCTGCGGCTTGGG 397
Db 322 AGCGCAGGAGGTGAGGATCAGCAGCGCGGCTCTGCGCGCAGCTGCTGCGGCTTGGG 381
QY 398 GCGCGCCCGCCCACTCTGATCCGGCTTGGGCTTGGAGCAGCAGCCCGAGCGCGCTGCAG 457
Db 382 GCGCGCCCGCCCACTCTGATCCGGCTTGGGCTTGGAGCAGCAGCCCGAGCGCGCTGCAG 441
QY 458 CGCAGCTCGCTCGGCTCTGCTCGCGCGCGGCTTGAACCTGCGGCTCAGCAGCCCGAGC 517
Db 442 CGCAGCTCGCTCGGCTCTGCTCGCGCGCGGCTTGAACCTGCGGCTCAGCAGCCCGAGC 501
QY 518 TTGTCCTCCGCGCGGCTCCCGCGGCTCAGCAGCCCGCGGCTTGGAGCGCGGCTTGGAGCGAGC 577
Db 502 TTGTCCTCCGCGCGGCTCCCGCGGCTCAGCAGCCCGCGGCTTGGAGCGCGGCTTGGAGCGAGC 561
QY 578 GCGCGCGCGGCGGATGCTGAGGAGCGCGGCTCAGCAGCAGCCCGAGCGCGGCTTGGAGCGCGG 637
Db 562 GCGCGCGCGGCGGATGCTGAGGAGCGCGGCTCAGCAGCAGCCCGAGCGCGGCTTGGAGCGCGG 621
QY 638 TGTGAGGTACTTGTGCTGGG 656
Db 622 TGTGAGGTACTTGTGCTGGG 640

RESULT 4
BI669134
LOCUS
DEFINITION
BI669134
603295564F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5314814 5',
mRNA sequence.
```



```
Query Match          51.5%; Score 504; DB 12; Length 570;
Best Local Similarity 100.0%; Pred. No. 1.9e-220;
Matches 504; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 476 TGTCTCGCGCGCCGCTTACCTGCGCGCTAGCAGCCAGCTTGTCCCGCGCGCTGC 535
Db      |||||||
QY 536 TGTCTCGCGCGCCGCTTACCTGCGCGCTAGCAGCCAGCTTGTCCCGCGCGCTGC 477
Db      |||||||
QY 536 CGCGCGCGCGCTTCCGACCCCGCGCGCTTACGACGACGCGCGCGCGCGCGATG 595
Db      |||||||
QY 476 CGCGCGCGCGCTTCCGACCCCGCGCGCTTACGACGACGCGCGCGCGCGCGATG 417
Db      |||||||
QY 596 CTGAGGAGCAGCGGAGACACCGAGCTGGACCCCGAGCTGTGAGTACTTGTCTGG 555
Db      |||||||
QY 416 CTGAGGAGCAGCGGAGACACCGAGCTGGACCCCGAGCTGTGAGTACTTGTCTGG 357
Db      |||||||
QY 656 GACGATTCTTGGGGAAGCGGAGCTCCGAGGGGTGACGCGCGCGCGCTCCGCG 715
Db      |||||||
QY 356 GACGATTCTTGGGGAAGCGGAGCTCCGAGGGGTGACGCGCGCGCGCTCCGCG 297
Db      |||||||
QY 716 GTGCCCGCGGACGACGATGCGGCTGTGAGCTGCCCCCTGAGGCGCGTGTGGGGCGCTGC 775
Db      |||||||
QY 296 GTGCCCGCGGACGACGATGCGGCTGTGAGCTGCCCCCTGAGGCGCGTGTGGGGCGCTGC 237
Db      |||||||
QY 776 TGCCTGTGAAGCCCTAGAGACCCCGCGCGCGCGAGTGCTGACGCGCGCTTGTGCCAC 835
Db      |||||||
QY 236 TGCCTGTGAAGCCCTAGAGACCCCGCGCGCGAGTGCTGACGCGCGCTTGTGCCAC 177
Db      |||||||
QY 836 CTGTGACACTGCGCGGATCCGCTGACCTGGACCCAGAGTCCCGCGCGCATCCGCG 895
Db      |||||||
QY 176 CTGTGACACTGCGCGGATCCGCTGACCTGGACCCAGAGTCCCGCGCGCATCCGCG 117
Db      |||||||
QY 896 ACCAGGACTGTCTCCCGCGGACGACGCTCCAGAGCACTTACCCTCCGCGCGAGCCCTCTC 955
Db      |||||||
QY 116 ACCAGGACTGTCTCCCGCGGACGACGCTCCAGAGCACTTACCCTCCGCGCGAGCCCTCTC 57
Db      |||||||
QY 956 ACCGAGGATCCCTACCCCTGCG 979
Db      |||||||
QY 56 ACCGAGGATCCCTACCCCTGCG 33
Db      |||||||

RESULT 6
BM876317/c
LOCUS      1j61a10.xl Human insulinoma Homo sapiens cDNA clone IMAGE:5635123
DEFINITION 3' similar to TR:Q9UH2 Q9UH2 PROSAS PRECURSOR. ;, mRNA sequence.
ACCESSION  BM876317
VERSION     BM876317.1 GI:19243983
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 589)
AUTHORS     Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
            M., Gibbons M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.,
            Jackson,Y. and Bowers,Y.
            Endocrine Pancreas Consortium
            Unpublished (2000)
            Other_ESTs: 1j61a10.y1
            Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. J. Ferrer In vivo mass-excised to
            pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington
```

University Genome Sequencing Center For information on obtaining a
clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
Seq primer: -400P from Gibco
High quality sequence stop: 415.

FEATURES
source
1..589
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5635123"
/tissue_type="Human insulinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pBluescript SK-; Site:1:
XhoI; Site:2: EcoRI; Constructed with lambda ZAPII system
(Stratagene) by Dr. J. Ferrer, in vivo mass-excised to
pBluescript SK- by Dr. H. Inoue following the Washington
University protocol
(http://genome.wustl.edu/est/lambda_protocol.shtml).
Please contact Hiroshi Inoue, MD/PhD for further
information on this library (Metabolism Division, Permutt
Laboratory, Washington University School of Medicine, Box
8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this
is a Washington University Pancreas EST project library."

BASE COUNT 78 a 173 c 237 g 100 t 1 others

Query Match 49.7%; Score 487; DB 14; Length 589;
Best Local Similarity 100.0%; Pred. No. 1.2e-212;
Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 493 GACCTTGCCTGCGCGCTTACGACCCAGCTTGTCCCGCGCGCTCCCGCGCGCTCCGA 552
Db |||||||
QY 529 GACCTTGCCTGCGCGCTTACGACCCAGCTTGTCCCGCGCGCTCCCGCGCGCTCCGA 470
Db |||||||
QY 553 CCCGCGCGCGCTTACGACGACGCGCGCGCGCGCGCGCTTGTCCCGCGCGCTCCCGCGCGCTCCGA 612
Db |||||||
QY 469 CCCGCGCGCGCTTACGACGACGCGCGCGCGCGCGCTTGTCCCGCGCGCTCCCGCGCGCTCCGA 410
Db |||||||
QY 613 GAGACACCGGAGTGAGACCCCGAGCTGTGAGTACTTGTGGGAGCGGATTTGTGGGGA 672
Db |||||||
QY 409 GAGACACCGGAGTGAGACCCCGAGCTGTGAGTACTTGTGGGAGCGGATTTGTGGGGA 350
Db |||||||
QY 673 ACCTGGGAGCTCCGAGGGGTGCGCGCGCGCGCTCCCGCGCGCTCCCGCGCGCTCCGA 732
Db |||||||
QY 349 AGCGGGAGCTCCGAGGGGTGCGCGCGCGCGCTCCCGCGCGCTCCCGCGCGCTCCGA 290
Db |||||||
QY 733 GTGGGCTCTGAGCTCCCGCTGAGGCGCGCTGTGGGGCGCGCTGTGGTGTGAACGCTTA 792
Db |||||||
QY 289 GTGGGCTCTGAGCTCCCGCTGAGGCGCGCTGTGGGGCGCGCTGTGGTGTGAACGCTTA 230
Db |||||||
QY 793 GAGACCCCGCGCGCGCTGAGTGTGACGCGCGCTTGTGGCACCCTGAGCACTGTGCCGA 852
Db |||||||
QY 229 GAGACCCCGCGCGCGCTGAGTGTGACGCGCGCTTGTGGCACCCTGAGCACTGTGCCGA 170
Db |||||||
QY 853 TCCCTGTGACCTGGGAGCCAGAGTGTCCCGCGCGCTCCCGCGCGCTCCCGCGCGCTCCGA 912
Db |||||||
QY 169 TCCCTGTGACCTGGGAGCCAGAGTGTCCCGCGCGCTCCCGCGCGCTCCCGCGCGCTCCGA 110
Db |||||||
QY 913 CCAGCAGCTCCAGAGCACTTACCCCGCGCGCTCCCGCGCGCTCCCGCGCGCTCCCGCGCGCTCCGA 972
Db |||||||
QY 109 CCAGCAGCTCCAGAGCACTTACCCCGCGCGCTCCCGCGCGCTCCCGCGCGCTCCCGCGCGCTCCGA 50
Db |||||||
QY 973 CCCTGGC 979
Db |||||||
QY 49 CCCTGGC 43
Db |||||||

RESULT 7
BM695925
LOCUS BM695925
DEFINITION UI-E-CL1-afa-a-11-0-UI.r1 UI-E-CL1 Homo sapiens cDNA clone
UI-E-CL1-afa-a-11-0-UI 5', mRNA sequence.
ACCESSION BM695925

599 bp mRNA linear EST 28-FEB-2002

```

VERSION      BM695925.1  GI:19009183
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 599)
AUTHORS      Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE        Normalization and subtraction: two approaches to facilitate gene
             discovery
JOURNAL      Genome Res. 6 (9), 791-806 (1996)
MEDLINE      97044477
COMMENT      Contact: Soares, MB
             Program for Rat Gene Discovery and Mapping
             University of Iowa
             451 Eckstein Medical Research Building Iowa City, IA 52242, USA
             Tel: 319 335 8250
             Fax: 319 335 9565
             Email: msoares@blue.weeg.uiowa.edu
             Tissue Procurement: Dr. Gregg Hageman
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Researchers may obtain clones from Research
             Genetics (www.resgen.com).
             The following repetitive elements were found in this cDNA
             sequence: 202-272, >GC-rich#Low_complexity
             Seq primer: M13 Reverse.
FEATURES     Location/Qualifiers
             1..599
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="UI-E-CL1-afa-a-11-0-UI"
             /clone_lib="UI-E-CL1"
             /tissue_type="human retina"
             /dev_stage="adult"
             /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
             /note="Organ: eye; Vector: p7T73-Pac (Pharmacia) with a
             modified polylinker; Site_1: EcoR I; Site_2: Not I;
             UI-E-CL1 is a normalized cDNA library containing the
             following tissue(s): retina. The library was constructed
             according to Bonaldo, Lennon and Soares, Genome Research,
             6:791-806, 1996. First strand cDNA synthesis was primed
             with an oligo-dT primer containing a Not I site. Double
             stranded cDNA was ligated to an EcoR I adaptor, digested
             with Not I, and cloned directionally into p7T73-Pac
             vector. The oligonucleotide used to prime the synthesis of
             first-strand cDNA contains a library tag sequence that is
             located between the Not I site and the (dr)18 tail. The
             sequence tag for this library is CCGCG. This library was
             created for the program, Gene Discovery in the Visual
             System, supported by National Eye Institute (NEI)."
```

BASE COUNT 78 a 236 c 207 g 78 t

ORIGIN

Query Match 49.4%; Score 484; DB 14; Length 599;
Best Local Similarity 99.8%; Pred. No. 2.9e-211;
Matches 534; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 321 GGACGGGGCGGGCCGAGCGGAGGAGGCTGAGATCAGCAGCGCGCTCTGGCGCA 380
Db 1 GGACGGGGCGGGCCGAGCGGAGGAGGCTGAGATCAGCAGCGCGCTCTGGCGCA 60
QY 381 GCTGCTGGGGTCTGGGGCGGGCCGCAACTCTGATCCGGCTCTGGGCTTGGACGACGA 440
Db 61 GCTGCTGGGGTCTGGGGCGGGCCGCAACTCTGATCCGGCTCTGGGCTTGGACGACGA 120
QY 441 CCGCAGCGGCTGACGCGAGCTCGTCGCGCTCTGCTCCGGCGCGGCTTGACCTGC 500
Db 121 CCGCAGCGGCTGACGCGAGCTCGTCGCGCTCTGCTCCGGCGCGGCTTGACCTGC 180
QY 501 CGCCCTAGCAGCCAGCTTGTCCCGGGCGGGTCCCGCGCGGGCGCTCCGACCCGGCC 560

```

```

Db 181 GCCTAGCAGCCAGCTGTCCCGCGCGCTCCCGCGCGGGCTCCGACCCCGGCC 240
QY 561 CCCGCTCTACGACAGCGCCCGCGGGCCCGGATGCTGAGGAGGCGAGCGAGACACC 620
Db 241 CCCGCTCTACGACAGCGCCCGCGGGCCCGGATGCTGAGGAGGCGAGCGAGACACC 300
QY 621 CGACGTGACACCCCGAGCTGTGAGTACTTCTGCTGGAGCGGATTCTTGGGGAAGCGGGA 680
Db 301 CGACGTGACACCCCGAGCTGTGAGTACTTCTGCTGGAGCGGATTCTTGGGGAAGCGGGA 360
QY 681 CTCCGAGGGGTGGAGCCCGCGCGCTCCGCGCGCTCCGCGCGGATGTGGGCTC 740
Db 361 CTCCGAGGGGTGGAGCCCGCGCGCTCCGCGCGCTCCGCGCGGATGTGGGCTC 420
QY 741 TGAGTGGCCCTGAGGGCGTGTGGGGCGCTGCTGGTGTGAAACCCCTAGAGACCC 800
Db 421 TGAGTGGCCCTGAGGGCGTGTGGGGCGCTGCTGGTGTGAAACCCCTAGAGACCC 480
QY 801 GGCCCGCCAGTGCCTGCACGCCCGCTTGTCCACCCCTGAGCACTGCCCGGATCC 855
Db 481 GGCGCCCGAGTGCCTGCACGCCCGCTTGTCCACCCCTGAGCACTGCCCGGATCC 535

```

RESULT 8

BM720023

LOCUS BM720023 537 bp mRNA linear EST 01-MAR-2002

DEFINITION UI-E-EJ0-ahu-d-07-0-UI.r1 UI-E-EJ0 Homo sapiens cDNA clone

ACCESSION BM720023

VERSION BM720023.1 GI:19038987

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 537)

AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 97044477

COMMENT Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

Tissue Procurement: Dr. Gregg Hageman

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Researchers may obtain clones from Research

Genetics (www.resgen.com).

The following repetitive elements were found in this cDNA

sequence: 202-272, >GC-rich#Low_complexity

Seq primer: M13 Reverse.

Location/Qualifiers

1..537

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="UI-E-CL1-afa-a-11-0-UI"

/clone_lib="UI-E-CL1"

/tissue_type="human retina"

/dev_stage="adult"

/lab_host="DH10B (Life Technologies) (T1 phage resistant)"

/note="Organ: eye; Vector: p7T73-Pac (Pharmacia) with a

modified polylinker; Site_1: EcoR I; Site_2: Not I;

UI-E-CL1 is a normalized cDNA library containing the

following tissue(s): retina. The library was constructed

according to Bonaldo, Lennon and Soares, Genome Research,

6:791-806, 1996. First strand cDNA synthesis was primed

with an oligo-dT primer containing a Not I site. Double

stranded cDNA was ligated to an EcoR I adaptor, digested

with Not I, and cloned directionally into p7T73-Pac

vector. The oligonucleotide used to prime the synthesis of

first-strand cDNA contains a library tag sequence that is

located between the Not I site and the (dr)18 tail. The

sequence tag for this library is CCGCG. This library was

created for the program, Gene Discovery in the Visual

System, supported by National Eye Institute (NEI)."

with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p7733-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tags for this library are: fetal eyes, AGAATCAGA; lens, CGATTAGGGA; eye anterior segment, AATCCGCAT; optic nerve, CCATTAGTG; retina, CCGCG; Retina Foveal and Macular, GTCC; RPE and Choroid, ACCTGA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

BASE COUNT 71 a 229 c 169 g 68 t
ORIGIN
Query Match 49.3%; Score 483; DB 14; Length 537;
Best Local Similarity 99.8%; Pred. No. 8.2e-211;
Matches 533; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 431 TGGACGAGGACCGGCGCTGCAGCGAGCTCGGTGCGGTCTGCTCGCGCCGCC 490
Db 4 TGGACGAGGACCGGCGCTGCAGCGAGCTCGGTGCGGTCTGCTCGCGCCGCC 63
QY 491 TTGACCTGCGCGCTAGCAGCCAGCTGTCCCGCGCCGTCGCCCGCGCGCTCC 550
Db 64 TTGACCTGCGCGCTCGCAGCGAGCTGTCCCGCGCGCTCGCGCGCGCGCTCC 123
QY 551 GACCCCGCGCGCTAGCAGCGAGCGCGCGCGCGCGCGCTGTGAGGAGGAGCGG 610
Db 124 GACCCCGCGCGCTACGACGAGCGCGCGCGCGCGCGCGCTGTGAGGAGGAGCGG 183
QY 611 ACGAGACCGGAGTGGACCGCGAGCTGTGAGGTACTGTGCGGGACGATTTTCGG 670
Db 184 ACGAGACCGGAGTGGACCGCGAGCTGTGAGGTACTGTGCGGGACGATTTTCGG 243
QY 671 GAAGCGGGACTCCGAGGGGTGCAGCGCGCGCGCGCTCCGCGCGCGCGACACG 730
Db 244 GAAGCGGGACTCCGAGGGGTGCAGCGCGCGCGCGCTCCGCGCGCGCGACACG 303
QY 731 ATGTGGGCTCTGAGTCCCGCTGAGGGCGTGTGGGGCGCTGCTGCGTGTGAAGGCC 790
Db 304 ATGTGGGCTCTGAGTCCCGCTGAGGGCGTGTGGGGCGCTGCTGCGTGTGAAGGCC 363
QY 791 TAGAGACCGCGCGCGCGAGTGCAGCGCGCGCTTTGCCACCGCTGAGCACTGCCG 850
Db 364 TAGAGACCGCGCGCGCGAGTGCAGCGCGCGCTTTGCCACCGCTGAGCACTGCCG 423
QY 851 GATCCCTGACACCTGGACCCAGAACTGCCCGCGCATCCGCGCACGAGACTGCTCC 910
Db 424 GATCCCTGACACCTGGACCCAGAACTGCCCGCGCATCCGCGCACGAGACTGCTCC 483
QY 911 CGCCAGCAGCTCCAGACAACTTACCCCGCGCGAGCCAGCCGCTCTCACCCGAGGA 964
Db 484 CGCCAGCAGCTCCAGACAACTTACCCCGCGCGAGCCAGCCGCTCTCACCCGAGGA 537

RESULT 9
BM504127/c
LOCUS
DEFINITION
h21e07.x1 Human insulinoma Homo sapiens cDNA 3', similar to
TF:Q9UHG2 Q9UHG2 PROSAS PRECURSOR. ; mRNA sequence.

ACCESSION
BM504127

VERSION
BM504127.1

KEYWORDS
GI:18666535

SOURCE
EST.

ORGANISM
human.

REFERENCE
Homo sapiens

AUTHORS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 593)

Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,

Lemishka,I., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S.,

Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,

Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas

M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.,
Jackson,Y. and Bowers,Y.
Endocrine Pancreas Consortium
Unpublished (2000)
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. J. Ferrer In vivo mass-excised to
pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington
University Genome Sequencing Center For information on obtaining a
clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
Seq primer: -40UP from Gibco
High quality sequence stop: 427.

FEATURES
Location/Qualifiers
1..593
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Human Insulinoma"
/tissue_type="insulinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: pancreas; Vector: pBluescript SK-; Site:1:
xhoI; Site:2: EcoRI; Constructed with lambda ZAPII system
(Stratagene) by Dr. J. Ferrer, In vivo mass-excised to
pBluescript SK- by Dr. H. Inoue following the Washington
University protocol
(http://genome.wustl.edu/est/lambda_protocol.shtml).
Please contact Hiroshi Inoue, MD/PhD for further
information on this library (Metabolism Division, Permutt
Laboratory, Washington University School of Medicine, Box
8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this
is a Washington University Pancreas EST project library."

BASE COUNT 80 a 190 c 248 g 74 t
ORIGIN
Query Match 48.8%; Score 478; DB 13; Length 593;
Best Local Similarity 100.0%; Pred. No. 1.6e-208;
Matches 478; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 476 TGCTCCGCGCGCTTGACCTGCGCGCTAGCAGCCAGCTGTGTCGCCGCGCGCTCC 535
Db 478 TGCTCCGCGCGCTTGACCTGCGCGCTAGCAGCCAGCTGTGTCGCCGCGCGCTCC 419
QY 536 CGCGCGCGCGCTCCGACCGCGCGCGCTCTACGACGAGCGCGCGCGCGCGGATG 595
Db 418 CGCGCGCGCGCTCCGACCGCGCGCGCTCTACGACGAGCGCGCGCGCGCGGATG 359
QY 596 CTGAGGAGGCGCGAGACACCGCGCGCTGGACCGCGAGCTGTGAGGTACTTGTCTGG 655
Db 358 CTGAGGAGGCGCGAGACACCGCGCGCTGGACCGCGAGCTGTGAGGTACTTGTCTGG 299
QY 656 GACGATTTCTGGGGAAAGCGCGGACTTCCGAGGGGGTGGCAGCGCGCGCGCGCTCCGCC 715
Db 298 GACGATTTCTGGGGAAAGCGCGGACTTCCGAGGGGGTGGCAGCGCGCGCGCGCTCCGCC 239
QY 716 GTCCCGCGGACCGAGTGGGTCTGAGCTGCCCCCTGAGGGCGGTCTGGGGCGGTCTG 775
Db 238 GTCCCGCGGACCGAGTGGGTCTGAGCTGCCCCCTGAGGGCGGTCTGGGGCGGTCTG 179
QY 776 TGGGTGTGAACGCTAGAGACCGCGCGCGCTGCGAGGTGCTGACGCGCGCTCTTTCGCAC 835
Db 178 TGGGTGTGAACGCTAGAGACCGCGCGCGCTGCGAGGTGCTGACGCGCGCTCTTTCGCAC 119
QY 836 CCTGAGCACTGCCCGGATCCCGTGCACCGCTGGGACCGCAGAGTGCCCGCGCGCTCCGCC 895
Db 118 CCTGAGCACTGCCCGGATCCCGTGCACCGCTGGGACCGCAGAGTGCCCGCGCGCTCCGCC 59
QY 896 ACAGGAGTCTGCCCGCGGACCGAGCTCCAGAGCAACTTACCGCGCGCGCGCGCGCTCC 953
|||||

```

Db 58 ACCAGGACTGCTCCCGCCGACAGCTCCAGACAACTTACCCCGCCAGCCAGCCCTC 1
RESULT 10
Bi711080/c 601 bp mRNA linear EST 11-MAR-2002
LOCUS i490d01.x1 Human insulinoma Homo sapiens cDNA clone IMAGE:5023128
DEFINITION 3' similar to TR:Q9UHG2 Q9UHG2 PROSAA5 PRECURSOR. ; mRNA sequence.
ACCESSION Bi711080
VERSION Bi711080
KEYWORDS EST.
SOURCE Bi711080.1 GI:15686775
human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 601)
AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
Lemishka,I., Scarce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
,M., Gibbons,M., McCann,R., Cole,R., Tsagarishvili,R., Williams,T.,
Jackson,Y. and Bowers,Y.
Endocrine Pancreas Consortium
Unpublished (2000)
Other ESTs: i490d01.y1
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@bioh.harvard.edu
Library was constructed by Dr. J. Ferrer In vivo mass-excised to
pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington
University Genome Sequencing Center For information on obtaining a
clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
Seq primer: -400P from Glibco
High quality sequence stop: 435.
FEATURES
source
1. 601
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cd_name="IMAGE:5023128"
/clone_lib="Human insulinoma"
/tissue_type="insulinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pBluescript SK-; Site:1:
XhoI; Site:2: EcoRI; Constructed with lambda ZAPII system
(Stratagene) by Dr. J. Ferrer, in vivo mass-excised to
pBluescript SK- by Dr. H. Inoue following the Washington
University protocol
(http://genome.wustl.edu/est/lambda_protocol.shtml).
Please contact Hiroshi Inoue, MD/PhD for further
information on this library (Metabolism Division, Permutt
Laboratory, Washington University School of Medicine, Box
8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this
is a Washington University Pancreas EST project library."
BASE COUNT 84 a 184 c 237 g 96 t
ORIGIN
Query Match 48.4%; Score 474; DB 13; Length 601;
Best Local Similarity 100.0%; Pred. No. 1.le-206;
Matches 474; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 506 TAGCAGCCAGCTTCTCCCGCCGCGCTCCCGCGCGCGCTCCGACCCCGCCCGG 565
Db 513 TAGCAGCCAGCTTCTCCCGCCGCGCTCCCGCGCGCGCTCCGACCCCGCCCGG 454
Qy 566 TCTACGACGAGCGCCCGCCGCGCGCTGCTGAGGAGCGAGCGGACGACACCCGACG 625
Db 453 TCTACGACGAGCGCCCGCCGCGCGCTGCTGAGGAGCGAGCGGACGACACCCGACG 394

```

```

Qy 626 TGGACCCCGAGCTGTGAGGTACTTGTGGGACGAGTCTTTCGGGGAAGCGCGGACTCG 685
Db 393 TGGACCCCGAGCTGTGAGGTACTTGTGGGACGAGTCTTTCGGGGAAGCGCGGACTCG 334
Qy 686 AGGGGGTGGCAGCCCGCGCGCTCCCGCGCGCGCGACACCATGTGGGCTCTGAGC 745
Db 333 AGGGGGTGGCAGCCCGCGCGCTCCCGCGCGCGCGACACCATGTGGGCTCTGAGC 274
Qy 746 TGCCCGCTGAGGGCGCTGTGGGGGCGCTGCTGCTGTGAAGCGCTAGAGACCCCGGCGC 805
Db 273 TGCCCGCTGAGGGCGCTGTGGGGGCGCTGCTGCTGTGAAGCGCTAGAGACCCCGGCGC 214
Qy 806 CCCAGTGCCTGCAGCGCGCTTCTTGGCACCTGAGCACTGCCCGGATCCCGTGCACCT 865
Db 213 CCCAGTGCCTGCAGCGCGCTTCTTGGCACCTGAGCACTGCCCGGATCCCGTGCACCT 154
Qy 866 GGGACCCAGAGTGCCTCCCGCGCGCTCCCGCGCGCGCTCCCGCGCGCGAGCTCCAG 925
Db 153 GGGACCCAGAGTGCCTCCCGCGCGCTCCCGCGCGCGCTCCCGCGCGAGCTCCAG 94
Qy 926 AGCAACTTACCCCGCGCGCGCTCTCACCGGAGTCCCTACCCCGCTGGC 979
Db 93 AGCAACTTACCCCGCGCGCGCTCTCACCGGAGTCCCTACCCCGCTGGC 40
RESULT 11
BM052768/c 515 bp mRNA linear EST 12-MAR-2002
LOCUS ie59g07.x2 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
DEFINITION cDNA clone IMAGE:5672268 3' similar to TR:Q9UHG2 Q9UHG2 PROSAA5
PRECURSOR. ; mRNA sequence.
ACCESSION BM052768
VERSION BM052768
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 515)
AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
Lemishka,I., Scarce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
,M., Gibbons,M., McCann,R., Cole,R., Tsagarishvili,R., Williams,T.,
Jackson,Y. and Bowers,Y.
Endocrine Pancreas Consortium
Unpublished (2000)
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@bioh.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Juliana Brown
(brown@fas.harvard.edu) This sequence now available from the IMAGE
consortium, for clone orders contact: info@image.llnl.gov
High quality sequence stop: 421.
FEATURES
source
1. 515
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5672268"
/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
/sex="Both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH10B"
/note="Organ: Pancreas; Vector: pSPORT1; Site:1: Not 1;
Site:2: Sal 1; Starting library constructed using

```

SuperScript Plasmid Library kit (Life Technologies). cDNA made by oligo-dT priming. Size-selected by column fractionation; average insert size 1.08 kb. Library was amplified once on solid support and plasmid DNA from library was prepared. The library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:791-806; 0.5 microgram single-stranded library plasmid DNA was mixed with 5 micrograms PCR product representing library inserts and hybridized to an EcoT of 20. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."

```
BASE COUNT      68 a  155 c  215 g   77 t
ORIGIN
Query Match      48.2%; Score 472; DB 13; Length 515;
Best Local Similarity 100.0%; Pred. No. 9e-206;
Matches 472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 508 GCAGCCAGCTGTGTCGCCGCGCGCTCCGCGCGCGGCTCCGACCCCGCGCGTC 567
Db 497 GCAGCCAGCTGTGTCGCCGCGCGCTCCGCGCGCGGCTCCGACCCCGCGCGTC 438
QY 568 TACGACGACGCGCGCGCGCGGATGCTGAGGAGCAGGCGAGACACCGAGCTG 627
Db 437 TACGACGACGCGCGCGCGCGGATGCTGAGGAGCAGGCGAGACACCGAGCTG 378
QY 628 GACCCGAGCTGTGAGGTACTTCTGGGACGGATTCTTCGGGAAGCGGACTCCGAG 687
Db 377 GACCCGAGCTGTGAGGTACTTCTGGGACGGATTCTTCGGGAAGCGGACTCCGAG 318
QY 688 GGGGTGCGACGCGCGCGCGCTCCGCGCGGCTCCGCGCGAGCAGATGCGGCTGAGCTG 747
Db 317 GGGGTGCGACGCGCGCGCGCTCCGCGCGGCTCCGCGCGAGCAGATGCGGCTGAGCTG 258
QY 748 CCCCTGAGGCGTGTGCGGGGCGCTGCTGCTGTGAAGCCCTAGAGACCCCGCGCGCC 807
Db 257 CCCCTGAGGCGTGTGCGGGGCGCTGCTGCTGTGAAGCCCTAGAGACCCCGCGCGCC 198
QY 808 CAGTGCTGACGCGCGCTTTTGCCACCTGTGACACTGCGCGGATCCCGTGACCCCTGG 867
Db 197 CAGTGCTGACGCGCGCTTTTGCCACCTGTGACACTGCGCGGATCCCGTGACCCCTGG 138
QY 868 GACCCAGAAGTGCCTCCGCGATCCCGCGCAGGAGCTGCTCCCGCGCAGACGTCAGAG 927
Db 137 GACCCAGAAGTGCCTCCGCGATCCCGCGCAGGAGCTGCTCCCGCGCAGACGTCAGAG 78
QY 928 CAACTTACCCGCGCAGCAGCGCTCTCACCCGAGGATCCCTACCCCTGCG 979
Db 77 CAACTTACCCGCGCAGCAGCGCTCTCACCCGAGGATCCCTACCCCTGCG 26
```

```
RESULT 12
BI789216/c
LOCUS
DEFINITION
  Ie53f09.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
  cDNA clone IMAGE:5670688 3' similar to TR:Q9UH2 Q9UH2 PROSARS
  PRECURSOR. ; mRNA sequence.
ACCESSION
  BI789216
VERSION
  BI789216.1 GI:15816941
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 518)
AUTHORS
  Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
  Lemishka,I., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
  Hillier,D., Marra,M., Pape,D., Wyllie,T., Martin,J., Blistain,A.,
  Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
  M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.,
  Jackson,Y. and Bowers,Y.
  Endocrine Pancreas Consortium
TITLE
```

JOURNAL COMMENT

Unpublished (2000)
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@bioh.p.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Juliana Brown
(brown@fas.harvard.edu) This sequence now available from the IMAGE
consortium, for clone orders contact: info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 435.

FEATURES

source

```
1..518
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5670688"
/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
/sex="Both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH10B"
/note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;
Site_2: Sal 1; Starting library constructed using
SuperScript plasmid library kit (Life Technologies). cDNA
made by oligo-dT priming. Size-selected by column
fractionation; average insert size 1.08 kb. Library was
amplified once on solid support and plasmid DNA from
library was prepared. The library DNA was normalized by
method #4 from Bonaldo, Lennon, and Soares 1996 Genome
Research 6:791-806; 0.5 microgram single-stranded library
plasmid DNA was mixed with 5 micrograms PCR product
representing library inserts and hybridized to an EcoT of
20. Single-stranded (unhybridized) plasmids were isolated
by hydroxyapatite chromatography and used to make this
library."
```

BASE COUNT 68 a 155 c 215 g 80 t
ORIGIN

Query Match 48.2%; Score 472; DB 13; Length 518;
Best Local Similarity 100.0%; Pred. No. 9.1e-206;
Matches 472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 508 GCAGCCAGCTGTGTCGCCGCGCGCTCCGCGCGCGGCTCCGACCCCGCGCGTC 567
Db 500 GCAGCCAGCTGTGTCGCCGCGCGCTCCGCGCGCGGCTCCGACCCCGCGCGTC 441
QY 568 TACGACGACGCGCGCGCGCGGATGCTGAGGAGCAGGCGAGACACCGAGCTG 627
Db 440 TACGACGACGCGCGCGCGCGGATGCTGAGGAGCAGGCGAGACACCGAGCTG 381
QY 628 GACCCGAGCTGTGAGGTACTTCTGGGACGGATTCTTCGGGAAGCGGACTCCGAG 687
Db 380 GACCCGAGCTGTGAGGTACTTCTGGGACGGATTCTTCGGGAAGCGGACTCCGAG 321
QY 688 GGGGTGCGACGCGCGCGCTCCGCGCGGCTCCGCGCGAGCAGATGCGGCTGAGCTG 747
Db 320 GGGGTGCGACGCGCGCGCTCCGCGCGGCTCCGCGCGAGCAGATGCGGCTGAGCTG 261
QY 748 CCCCTGAGGCGCTGTGCGGGGCGCTGCTGCTGTGAAGCCCTAGAGACCCCGCGCGCC 807
Db 260 CCCCTGAGGCGCTGTGCGGGGCGCTGCTGCTGTGAAGCCCTAGAGACCCCGCGCGCC 201
QY 808 CAGTGCTGCGACGCGCGCTTCTGCGACCTGAGACTGCGCGGATCCCGTGACACCTGG 867
Db 200 CAGTGCTGCGACGCGCGCTTCTGCGACCTGAGACTGCGCGGATCCCGTGACACCTGG 141
QY 868 GACCCAGAAGTGCCTCCGCGATCCCGCGCAGGAGCTGCTCCCGCGCAGCAGCTCCAGAG 927
```

```
Db 140 GACCCAGAGTGCCTCCGACAGGACTGCTCCCGCCACGACGTCACAG 81
QY 928 CAACCTACCCGCGCAGCAGCCCTCTACCCGAGGATCCCTACCCCTGGC 979
|||||
Db 80 CAACCTACCCGCGCAGCAGCCCTCTACCCGAGGATCCCTACCCCTGGC 29
|||||

RESULT 13
BF724971/c
LOCUS
DEFINITION
  541 bp mRNA linear EST 05-JAN-2001
  sapiens cDNA clone bx10f06 3', mRNA sequence.
ACCESSION
  BF724971
VERSION
  BF724971.1 GI:12040882
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 541)
REFERENCE
  Wistow,G.J., Bernstein,S., Behal,A. and Smith,D.
  NEIBANK: EST analysis and bioinformatics for ocular genomics
  Invest. Ophthalmol. Vis. Sci. 41, (2000) In press
  Contact: Wistow G
  Section on Molecular Structure and Function
  National Eye Institute
  6/331, NIH, Bethesda, MD 20892-2740, USA
  Tel: 301 402 3452
  Fax: 301 496 0078
  Email: graeme@helix.nih.gov
  Plate: 10 row: f column: 06
  Seq primer: -21M13 forward primer (ABI).
  Location/Qualifiers
    1..541
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="bx10f06"
      /clone_lib="Human Iris cDNA (Un-normalized, unamplified):
      BX"
      /tissue_type="Iris"
      /dev_stage="Adult"
      /lab_host="EMDH10B"
      /note="Organ: Eye; Vector: pCMVSPORT6; Post-mortem iris
      tissue was pooled from 10 individuals ranging in age from
      4-80 years and RNA was extracted. From this pooled sample
      an aliquot of 60ug of total RNA yielded 2.17ug of mRNA. A
      directionally cloned cDNA library in the pCMVSPORT6 vector
      was constructed at Life Technologies, essentially
      following the protocols of the Superscript Plasmid System
      full details of which are contained in the manufacturer's
      instruction manual (http://www.lifetech.com/). First
      strand synthesis was carried out using a Not I
      primer-adaptor [5'-pGACTAGTCTCTAGATCGGCGGCCCT(T)15-3',
      1. Not I/blunt end inserts were cloned into the Not I/EcoR
      V sites in the vector. EST analysis was performed on the
      unamplified library at the NIH Intramural Sequencing
      Center (NISC)."]
BASE COUNT
  69 a 159 c 222 g 91 t
ORIGIN

Query Match
  48.2%; Score 472; DB 12; Length 541;
  Best Local Similarity 100.0%; Pred. No. 9.1e-206;
  Matches 472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 508 GCAGCCAGCTTGTCCTCCGCGCCCTCCCGCGGCGCTCCGACCCGCGCCGTC 567
|||||
Db 512 GCAGCCAGCTTGTCCTCCGCGCCCTCCCGCGGCGCTCCGACCCGCGCCGTC 453
|||||
QY 568 TACGACGACGCGCCGCGGCGCGGATGCTGAGGAGGCGAGCGACACCCGACGTC 627
|||||
Db 452 TACGACGACGCGCCGCGGCGCGGATGCTGAGGAGGCGAGCGACACCCGACGTC 393
|||||
QY 628 GACCCGAGCTGTTGAGTACTTGTGGGACGATCTTGGGGAGCGCGGACTCCGAG 687
```

```
Db 392 GACCCGAGCTGTTGAGTGACTTGTGGACGGAATCTTTGCGGGAAGCGGACTCCGAG 333
QY 688 GGGGTGGCAGCCCGCGCGCTCGCGCGTGCAGGACAGATGTGGGCTCTGAGCTG 747
|||||
Db 332 GGGGTGGCAGCCCGCGCGCTCGCGCGTGCAGGACAGATGTGGGCTCTGAGCTG 273
|||||
QY 748 CCCCTGAGGGCGTGTCTGGGGCGTGTGCTGTGAACGCTAGAGACCCCGCGGCC 807
|||||
Db 272 CCCCTGAGGGCGTGTCTGGGGCGTGTGCTGTGAACGCTAGAGACCCCGCGGCC 213
|||||
QY 808 CAGGTGCTGCAGCGCGCTCTTGCACCCCTGAGCACTGCCCGGATCCGCTGACCCCTGG 867
|||||
Db 212 CAGGTGCTGCAGCGCGCTCTTGCACCCCTGAGCACTGCCCGGATCCGCTGACCCCTGG 153
|||||
QY 868 GACCCAGAGTGCCTCCCGCATCCCGCCACAGGACTGCTCCCGCCACGACGTCACAG 927
|||||
Db 152 GACCCAGAGTGCCTCCCGCATCCCGCCACAGGACTGCTCCCGCCACGACGTCACAG 93
|||||
QY 928 CAACCTACCCGCGCAGCAGCCCTCTCACCGGAGGATCCCTACCCCTGGC 979
|||||
Db 92 CAACCTACCCGCGCAGCAGCCCTCTCACCGGAGGATCCCTACCCCTGGC 41
|||||

RESULT 14
BM142527/c
LOCUS
DEFINITION
  562 bp mRNA linear EST 12-MAR-2002
  if35d05.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
  cDNA clone IMAGE:5678265 3' similar to TR:09UH2 Q9UHG2 PROSAA5
  PRECURSOR. ; mRNA sequence.
ACCESSION
  BM142527
VERSION
  BM142527.1 GI:17152594
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 562)
REFERENCE
  Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
  Lemishka,I., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
  Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
  Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
  ,M., Gibbons,M., McCann,R., Cole,R., Tsagarelshvili,R., Williams,T.
  , Jackson,Y. and Bowers,Y.
  Endocrine Pancreas Consortium
  Other ESTs: if35d05.y1
  Unpublished (2000)
  Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
  Endocrine Pancreas Consortium
  Harvard University, Howard Hughes Medical Institute
  Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
  MA 02138
  Tel: 617-495-1812
  Fax: 617-495-8557
  Email: dmelton@iohph.harvard.edu
  Library was constructed by Dr. Douglas Melton DNA sequencing by:
  Washington University Genome Sequencing Center For information on
  obtaining a clone please contact: Juliana Brown
  (brown@fas.harvard.edu) This sequence now available from the IMAGE
  consortium, for clone orders contact: info@image.llnl.gov
  Seq primer: -400P from GIBCO
  High quality sequence stop: 426.
  Location/Qualifiers
    1..562
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="IMAGE:5678265"
      /clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
      /sex="Both"
      /tissue_type="Islets of Langerhans"
      /dev_stage="Adult"
      /lab_host="DH10B"
      /note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;
```

Site.2: Sal 1: Starting library constructed using Superscript Plasmid Library kit (Life Technologies). cDNA made by oligo-dr priming. Size-selected by column fractionation; average insert size 1.08 kb. Library was amplified once on solid support and plasmid DNA from method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:791-806; 0.5 microgram single-stranded library plasmid DNA was mixed with 5 micrograms PCR product representing library inserts and hybridized to an EcoT of 20. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."

BASE COUNT 77 a 171 c 235 g 79 t
ORIGIN

Query Match 48.2%; Score 472; DB 13; Length 562;
Best Local Similarity 100.0%; Pred. No. 9.1e-206;
Matches 472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 508 GCAGCCAGCTGTGTCGGCGCGCGTCCCGCGCGGCTCGAGACCCCGCGCGTCC 567
DB 496 GCAGCCAGCTGTGTCGGCGCGCGTCCCGCGCGGCTCGAGACCCCGCGCGTCC 437
QY 568 TACGACGACGCGCGCGCGCGGATGCTGAGGAGCGGAGGAGACCGGAGCTG 627
DB 436 TACGACGACGCGCGCGCGCGGATGCTGAGGAGCGGAGGAGACCGGAGCTG 377
QY 628 GACCCGAGCTGTGAGTACTTCTGGGAGCGGATCTTGGGAGCGGAGCTCCGAG 687
DB 376 GACCCGAGCTGTGAGTACTTCTGGGAGCGGATCTTGGGAGCGGAGCTCCGAG 317
QY 688 GGGGTGCGAGCGCGCGCGCGCTCCCGCGCGGAGGAGGAGGAGGAGGAGGAGGAG 747
DB 316 GGGGTGCGAGCGCGCGCGCGCTCCCGCGCGGAGGAGGAGGAGGAGGAGGAGGAG 257
QY 748 CCCCTGAGGCGCTGTGGGGCGCGCTGCTGGGTGTAAGACCCCTAGAGACCCCGCGCGCC 807
DB 256 CCCCTGAGGCGCTGTGGGGCGCGCTGCTGGGTGTAAGACCCCTAGAGACCCCGCGCGCC 197
QY 808 CAGGTGCTGCAGCGCGCGCTTGTCCACCTGAGCACTGCGCGGATCCCGTGCACCTCG 867
DB 196 CAGGTGCTGCAGCGCGCGCTTGTCCACCTGAGCACTGCGCGGATCCCGTGCACCTCG 137
QY 868 GACCCAGAAGTGCCCGCGCATCCCGACAGGAGTCTCCCGCGGAGGAGGAGGAGGAGGAG 927
DB 136 GACCCAGAAGTGCCCGCGCATCCCGACAGGAGTCTCCCGCGGAGGAGGAGGAGGAGGAG 77
QY 928 CAACCTACCCCGGCGAGCGCGCTCTACCCGAGGATCCCTACCCCTGCG 979
DB 76 CAACCTACCCCGGCGAGCGCGCTCTACCCGAGGATCCCTACCCCTGCG 25

RESULT 15
BI789206/c
LOCUS BI789206 569 bp mRNA linear EST 12-MAR-2002
DEFINITION ie3805.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
cDNA clone IMAGE:5670632 3' similar to TR:Q9UH2 Q9UH2 PROSAS
PRECURSOR. ; mRNA sequence.
ACCESSION BI789206
VERSION BI789206.1 GI:15816931
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 569)
REFERENCE
AUTHORS Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,
Lemishka, I., Scearce, M., Brestelli, J., Gradwohl, G., Clifton, S.,
Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistain, A.,
Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas
M., Gibbons, M., McCann, R., Cole, R., Tsagareishvili, R., Williams, T.,
Jackson, Y. and Bowers, Y.

TITLE
JOURNAL
COMMENT

Endocrine Pancreas Consortium
Unpublished (2000)
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@bioh.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center for information on
obtaining a clone please contact: Juliana Brown
(brown@fas.harvard.edu) This sequence now available from the IMAGE
consortium, for clone orders contact: info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 427.

FEATURES
source

1..569
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5670632"
/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
/sex="Both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH108"
/note="Organ: Pancreas; Vector: pSPORT1; Site.1: Not 1;
Site.2: Sal 1; Starting library constructed using
Superscript Plasmid Library kit (Life Technologies). cDNA
made by oligo-dr priming. Size-selected by column
fractionation; average insert size 1.08 kb. Library was
amplified once on solid support and plasmid DNA from
method #4 from Bonaldo, Lennon, and Soares 1996 Genome
Research 6:791-806; 0.5 microgram single-stranded library
plasmid DNA was mixed with 5 micrograms PCR product
representing library inserts and hybridized to an EcoT of
20. Single-stranded (unhybridized) plasmids were isolated
by hydroxyapatite chromatography and used to make this
library."

BASE COUNT 77 a 170 c 238 g 83 t 1 others
ORIGIN

Query Match 48.2%; Score 472; DB 13; Length 569;
Best Local Similarity 100.0%; Pred. No. 9.1e-206;
Matches 472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 508 GCAGCCAGCTGTGTCGGCGCGCGTCCCGCGCGGCTCGAGACCCCGCGCGTCC 567
DB 500 GCAGCCAGCTGTGTCGGCGCGCGTCCCGCGCGGCTCGAGACCCCGCGCGTCC 441
QY 568 TACGACGACGCGCGCGCGCGGATGCTGAGGAGCGGAGGAGGAGGAGGAGGAGGAG 627
DB 440 TACGACGACGCGCGCGCGCGGATGCTGAGGAGCGGAGGAGGAGGAGGAGGAGGAG 381
QY 628 GACCCGAGCTGTGAGTACTTCTGGGAGCGGATCTTGGGAGGAGGAGGAGGAGGAG 687
DB 380 GACCCGAGCTGTGAGTACTTCTGGGAGCGGATCTTGGGAGGAGGAGGAGGAGGAG 321
QY 688 GGGGTGCGAGCGCGCGCGCGCTCCCGCGCGGAGGAGGAGGAGGAGGAGGAGGAG 747
DB 320 GGGGTGCGAGCGCGCGCGCGCTCCCGCGCGGAGGAGGAGGAGGAGGAGGAGGAG 261
QY 748 CCCCTGAGGCGCTGTGGGGCGCGCTGCTGGGTGTAAGACCCCTAGAGACCCCGCGCGCC 807
DB 260 CCCCTGAGGCGCTGTGGGGCGCGCTGCTGGGTGTAAGACCCCTAGAGACCCCGCGCGCC 201
QY 808 CAGGTGCTGCAGCGCGCGCTTGTGCCACCTGAGCACTGCGCGGATCCCGTGCACCTCG 867
DB 200 CAGGTGCTGCAGCGCGCGCTTGTGCCACCTGAGCACTGCGCGGATCCCGTGCACCTCG 141
QY 868 GACCCAGAAGTGCCCGCGCATCCCGGAGGAGTGTCTCCCGCGGAGGAGGAGGAGGAG 927

|||||
Db 140 GACCCAGAGTCCCCCGCCATCCCGCCACCCAGGACTGCTCCCCCGCCAGCACGTCCAGAG 81
QY 928 CAACTTACCCCGCCAGCCAGCCCTCTCACCCGAGGATCCCTACCCCTGGC 979
|||||
Db 80 CAACTTACCCCGCCAGCCAGCCCTCTCACCCGAGGATCCCTACCCCTGGC 29

Search completed: April 12, 2003, 20:46:19
Job time : 1608 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 12, 2003, 20:46:29 ; Search time 152 Seconds
(without alignments)
5649.653 Million cell updates/sec

Title: US-09-803-589-5
Perfect score: 979
Sequence: 1 gaattcgacagcagcagc.....gagatccctaccctccctggc 979

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 593429 seqs, 438583890 residues

Word size : 15

Total number of hits satisfying chosen parameters: 3821

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	979	100.0	979	10	US-09-803-589-5
2	583	59.6	1050	9	US-09-984-271-58
3	226	23.1	301	10	US-09-880-107-1701
c	32	3.3	349	10	US-09-960-352-71
4	32	3.3	1027	10	US-09-803-589-13
5	32	3.3	1027	10	US-09-803-589-13
6	25	2.6	342	10	US-09-983-965-3684
7	20	2.0	372	10	US-09-983-965-2663
8	19	1.9	349	9	US-09-854-133-589
9	19	1.9	349	10	US-09-833-790-259
c	10	1.9	380	10	US-09-917-800A-799
c	11	1.9	424	10	US-09-983-965-4193
12	19	1.9	464	10	US-09-833-790-30
c	13	1.9	476	10	US-09-880-107-2439
14	19	1.9	796	9	US-09-981-876-24
15	19	1.9	796	9	US-09-148-545-24
c	16	1.9	942	10	US-09-754-949-1
17	19	1.9	972	9	US-09-920-923-33
18	19	1.9	1210	9	US-09-905-291A-126
19	19	1.9	1210	9	US-09-902-853-126

20	19	1.9	1210	9	US-09-907-824-126	Sequence 126, App
21	19	1.9	1210	9	US-09-907-841-126	Sequence 126, App
22	19	1.9	1210	9	US-09-904-011-126	Sequence 126, App
23	19	1.9	1210	9	US-10-028-072-311	Sequence 311, App
24	19	1.9	1210	9	US-09-906-742-126	Sequence 126, App
25	19	1.9	1210	9	US-10-121-049-311	Sequence 311, App
26	19	1.9	1210	9	US-10-123-904-311	Sequence 311, App
27	19	1.9	1210	9	US-10-140-470-311	Sequence 311, App
28	19	1.9	1210	9	US-09-906-838-126	Sequence 126, App
29	19	1.9	1210	9	US-09-907-613-126	Sequence 126, App
30	19	1.9	1210	9	US-09-907-942-126	Sequence 126, App
31	19	1.9	1210	9	US-10-175-746-311	Sequence 311, App
32	19	1.9	1210	9	US-10-176-918-311	Sequence 311, App
33	19	1.9	1210	9	US-10-137-865-311	Sequence 311, App
34	19	1.9	1210	9	US-10-140-474-311	Sequence 311, App
35	19	1.9	1210	9	US-10-143-114-311	Sequence 311, App
36	19	1.9	1210	9	US-09-904-820-126	Sequence 126, App
37	19	1.9	1210	9	US-09-904-859-126	Sequence 126, App
38	19	1.9	1210	9	US-09-909-204-126	Sequence 126, App
39	19	1.9	1210	9	US-10-142-431-311	Sequence 311, App
40	19	1.9	1210	9	US-10-143-114-311	Sequence 311, App
41	19	1.9	1210	9	US-09-904-786-126	Sequence 126, App
42	19	1.9	1210	9	US-09-906-646-126	Sequence 126, App
43	19	1.9	1210	9	US-09-906-700-126	Sequence 126, App
44	19	1.9	1210	9	US-10-140-002-311	Sequence 311, App
45	19	1.9	1210	9	US-09-902-903-126	Sequence 126, App

ALIGNMENTS

RESULT 1
US-09-803-589-5
; Sequence 5, Application US/09803589
; Patent No. US20020112251A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Goodearl, Andrew D.J.
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC AND OTHER
; TITLE OF INVENTION: USES
; FILE REFERENCE: 07334-325001
; CURRENT APPLICATION NUMBER: US/09/803,589
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: US 09/128,709
; PRIOR FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: US 60/054,645
; PRIOR FILING DATE: 1997-08-04
; PRIOR APPLICATION NUMBER: US 09/130,491
; PRIOR FILING DATE: 1998-08-06
; PRIOR APPLICATION NUMBER: US 60/054,966
; PRIOR FILING DATE: 1997-08-06
; PRIOR APPLICATION NUMBER: US 60/058,108
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: US 09/388,280
; PRIOR FILING DATE: 1999-09-01
; PRIOR APPLICATION NUMBER: US 09/388,279
; PRIOR FILING DATE: 1999-09-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 979
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (58)...(837)
US-09-803-589-5

Query Match 100.0%; Score 979; DB 10; Length 979;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 979; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 3684
; LENGTH: 342
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 43-LIB3058-043-Q1-K1-C12
US-09-983-965-3684

Query Match 2.6%; Score 25; DB 10; Length 342;
Best Local Similarity 100.0%; Pred. No. 0.0051;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 GAGGATCAGCAGCGCGCTCTGG 376
|||||
Db 4 GAGGATCAGCAGCGCGCTCTGG 28

RESULT 7

US-09-983-965-2663
; Sequence 2663, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Machialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 2663
; LENGTH: 372
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 10-LIB3058-043-Q1-K1-C5
US-09-983-965-2663

Query Match 2.0%; Score 20; DB 10; Length 372;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 508 GCAGCCAGCTGTCCCGC 527
|||||
Db 127 GCAGCCAGCTGTCCCGC 146

RESULT 8

US-09-854-133-589
; Sequence 589, Application US/09854133
; Publication No. US20020183499A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Henderson, Robert A.
; APPLICANT: Benson, Darin R.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 210121.475C10
; CURRENT APPLICATION NUMBER: US/09/854,133
; CURRENT FILING DATE: 2001-05-11

; NUMBER OF SEQ ID NOS: 735
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 589
; LENGTH: 349
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-854-133-589

Query Match 1.9%; Score 19; DB 9; Length 349;
Best Local Similarity 100.0%; Pred. No. 5.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCGCGCAGGCGCCAG 19
|||||
Db 1 GAATTCGCGCAGGCGCCAG 19

RESULT 9

US-09-833-790-259
; Sequence 259, Application US/09833790
; Patent No. US20020068288A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Wang, Tongtong
; APPLICANT: Secrist, Heather
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Indirias, Carol Y.
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.512
; CURRENT APPLICATION NUMBER: US/09/833,790
; CURRENT FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 259
; LENGTH: 349
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-833-790-259

Query Match 1.9%; Score 19; DB 10; Length 349;
Best Local Similarity 100.0%; Pred. No. 5.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCGCGCAGGCGCCAG 19
|||||
Db 1 GAATTCGCGCAGGCGCCAG 19

RESULT 10

US-09-917-800A-799/c
; Sequence 799, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castile, Arthur
; APPLICANT: Etashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15

```
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 799
; LENGTH: 380
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 A1136514
US-09-917-800A-799

Query Match      1.9%; Score 19; DB 10; Length 380;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AATTCGGCAGCGGCCAGC 20
Db 380 AATTCGGCAGCGGCCAGC 362

RESULT 11
US-09-983-965-4193/c
; Sequence 4193, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; PRIOR FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1998-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 4193
; LENGTH: 424
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 59-LIB3058-051-Q1-K1-G4
US-09-983-965-4193

Query Match      1.9%; Score 19; DB 10; Length 424;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 742 GAGCTGCCCTGAGGGCG 760
Db 108 GAGCTGCCCTGAGGGCG 90

RESULT 12
US-09-833-790-30
; Sequence 30, Application US/09833790
; Patent No. US20020068288A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Wang, TongLong
; APPLICANT: Secrist, Heather
```

```
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Indirias, Carol Y.
; APPLICANT: Fan, Lihun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.512
; CURRENT APPLICATION NUMBER: US/09/833,790
; CURRENT FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 464
; TYPE: DNA
; ORGANISM: Homo sapien
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 N20198
US-09-833-790-30

Query Match      1.9%; Score 19; DB 10; Length 464;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCGGCAGCGGCCAG 19
Db 1 GAATTCGGCAGCGGCCAG 19

RESULT 13
US-09-880-107-2439/c
; Sequence 2439, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2439
; LENGTH: 476
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 N20198
; NAME/KEY: unsure
; LOCATION: (1)..(476)
; OTHER INFORMATION: n = a o c o r g o t
US-09-880-107-2439

Query Match      1.9%; Score 19; DB 10; Length 476;
Best Local Similarity 100.0%; Pred. No. 4.8;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 TTTTGGTGTCTGCTGCT 125
Db 445 TTTTGGTGTCTGCTGCT 427

RESULT 14
US-09-981-876-24
; Sequence 24, Application US/09981876
; Patent No. US20020164669A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 70 Human Secreted Proteins
; FILE REFERENCE: P2001P1
; CURRENT APPLICATION NUMBER: US/09/981,876
```



```

; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,590
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,594
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,589
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,593
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,614
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/043,578
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/043,576
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/047,501
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/043,670
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/056,632
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,664
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,876
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,881
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,909
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,875
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,862
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,887
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,908
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/048,964
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 60/057,650
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: 60/056,884
; PRIOR FILING DATE: 1997-08-22
; NUMBER OF SEQ ID NOS: 280
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 796

Query Match      1.9%; Score 19; DB 9; Length 796;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 113 TGCTGCTGCTGCTCGGCGCT 131
      |||||
Db 62 TGCTGCTGCTGCTCGGCGCT 80

RESULT 15
US-09-148-545-24
; Sequence 24, Application US/09148545
; Publication No. US20030027132A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 70 Human Secreted Proteins
; FILE REFERENCE: P2001P1
; CURRENT APPLICATION NUMBER: US/09/148,545
; EARLIER FILING DATE: 1998-09-04
; EARLIER APPLICATION NUMBER: PCT/US98/04482
; EARLIER FILING DATE: 1998-03-06
; EARLIER APPLICATION NUMBER: 60/040,162
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,333
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/043,621
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,161
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,626
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,334
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,336
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,163
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/047,615
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,600
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,597
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,502
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,633
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,583
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,617
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,618
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,503
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,592
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,581
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,584
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,500
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,587
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,492
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,598
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,613
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,582
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,596
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,612
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,632
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,601
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,580
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,568
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,314
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,569
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,311
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,671
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,674
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,669
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,312
```

; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,313
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,672
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,315
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/048,974
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/056,886
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,877
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,889
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,893
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,630
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,878
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,662
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,872
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,882
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,637
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,903
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,888
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,879
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,880
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,894
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,911
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,636
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,874
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,910
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,864
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,631
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,845
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,892
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/047,595
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/057,761
; EARLIER FILING DATE: 05-Sep-1997
; EARLIER APPLICATION NUMBER: 60/047,599
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,588
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,585
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,586
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,590
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,594
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,589
; EARLIER FILING DATE: 1997-05-23

; EARLIER APPLICATION NUMBER: 60/047,593
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,614
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,578
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,576
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/047,501
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,670
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/056,632
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,664
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,876
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,881
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,909
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,875
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,862
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,887
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,908
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/048,964
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/057,650
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/056,884
; EARLIER FILING DATE: 1997-08-22
; NUMBER OF SEQ ID NOS: 280
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 796

Query Match 1.98; Score 19; DB 9; Length 796;
Best Local Similarity 100.0%; Pred. No. 4.3;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 113 TGCTGCTGCTGCTGCGCCT 131
Db 62 TGCTGCTGCTGCTGCGCCT 80

Search completed: April 12, 2003, 21:41:32
Job time : 162 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2003, 14:15:09 ; Search time 18 Seconds
(without alignments)
1388.608 Million cell updates/sec

Title: US-09-803-589-6
Perfect score: 1319
Sequence: 1 MAGSPLWGPRAGVGGLVL.....RVKRLTPAPQVPARRLLPP 260

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	146	11.1	913	2 B97369	translation initia
2	146	11.1	913	2 AB2587	translation initia
3	134	10.2	553	2 C75318	hypothetical prote
4	130.5	9.9	589	2 G87485	hypothetical prote
5	127.5	9.7	460	2 T33110	hypothetical prote
6	124.5	9.4	815	2 B56708	extracellular sign
7	123	9.3	801	2 T29018	hypothetical prote
8	123	9.3	1238	2 T03465	probable exonuclea
9	121.5	9.2	590	1 T35297	probable dihydroli
10	119.5	9.1	422	2 S32357	glial growth facto
11	119.5	9.1	1460	1 EDBE1F	immediate-early pr
12	118.5	9.0	1298	1 EDBE75	immediate-early pr
13	118	8.9	1446	1 A45344	gag polyprotein
14	117	8.9	643	1 FOLJLK	Xin protein, stage
15	116.5	8.8	1677	2 T14267	ETS2 repressor fac
16	116	8.8	548	2 S59133	probable ATP-bindl
17	115.5	8.8	998	2 T35745	probable multi-dom
18	115	8.7	2240	2 T37057	TonB-dependent rec
19	114	8.6	976	2 G87389	precorrin-3 methyl
20	113.5	8.6	559	2 F83283	protein kinase, tr
21	113.5	8.6	1089	2 T36663	UL36 protein - hum
22	113.5	8.6	3164	1 WMBEH6	cycH protein - Par
23	112.5	8.5	413	2 S61305	progesterone recep
24	112.5	8.5	933	1 QRHDP	smc protein (impor
25	112	8.5	817	2 T03852	GTP-binding regula
26	112	8.5	1147	2 D87295	conserved hypothet
27	111	8.4	846	2 S52418	MHC class III hist
28	111	8.4	1607	2 G87259	
29	111	8.4	1870	2 S37671	

30	111	8.4	1872	2 S36152	MHC class III hist
31	111	8.4	2142	2 B35098	MHC class III hist
32	110.5	8.4	1151	2 T18535	high molecular mas
33	110	8.3	744	2 T35192	probable ABC trans
34	109	8.3	681	2 A83455	DNA polymerase sub
35	108.5	8.2	3413	2 T17457	rifamycin polyketi
36	108.5	8.2	7463	2 T36248	CDA peptidase synth
37	108	8.2	327	2 S30405	hypothetical prote
38	108	8.2	929	2 T35683	flsk homolog - Str
39	107.5	8.2	382	2 T35709	hypothetical prote
40	107.5	8.2	637	2 A75342	hypothetical prote
41	107.5	8.2	1791	2 T02345	hypothetical prote
42	107	8.1	676	1 EDBE22	immediate-early pr
43	107	8.1	881	2 T03461	methyl-accepting c
44	107	8.1	954	1 S68178	mixed-lineage prot
45	107	8.1	2241	2 T02857	conserved hypothet

ALIGNMENTS

RESULT 1
B97369

translation initiation factor IF-2 [imported] - Agrobacterium tumefaciens (strain C58)
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C:Accession: B97369
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldm
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz,
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
A:Reference number: A97359; PMID:11743194
A:Accession: B97369
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-913 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK85907.1; PID:g15154954; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C_131
A:Map position: circular chromosome
C:Superfamily: translation initiation factor IF-2; translation elongation factor Tu h

Query Match	11.1%	Score 146;	DB 2;	Length 913;
Best Local Similarity	25.2%;	Pred. No. 0.034;		
Matches	65;	Conservative	34;	Mismatches 79;
				Indels 80; Gaps 12;
QY	29	PPALCARP-----VKPRGL-----SAAPPLAETGAPRRFRSVPRGE-----AACAV	72	
Db	72	PAPVAARPOAPQPRHQPGGQQQPRGSSQSQRRSSAPQQRQADRPNGVHDLUSAGEM	131	
QY	73	QELARALAHLEA-----ERQERARAEQAEQDQARVLAQLLRVWGAPRNSDPALGLD	126	
Db	132	EARRA---LMEAQARDVVEAKQRAEDEARRKVEEQRIAAE--KMEAANRAAEAAAK	186	
QY	127	DDPAPRAQLARALLRLDPAALAAQLVPAPVPAALPRPPVYDDGPGAPDAAEAGDE	186	
Db	187	VAASQPAAEV-----RAEPASERPAAPRTDARPO-----SAAAAPRSAPA---	231	
QY	187	TPVDVDPPELLRYLLGRILAGSADSEGVAPRRLRAADHD-----VGSLEPPEGVLCAL	239	
Db	232	TPD-----AAAPGRRTTGGDEDDRGAVRGSSLPARG-----	264	
QY	240	LRVKRLTPAPQVPARRL	257	
Db	265	---KVVPAPAKPAARL	278	

RESULT 2
AB2587

translation initiation factor IF-2 [imported] - Agrobacterium tumefaciens (strain C58)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
C:Accession: AB2587

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, Y.; Wang, Y.; Zhang, S.; Karp, P.; Romero, P.; Grant, C.; Guenther, D.; Kutysan, T.; Levy, R.; Li, M.; McClellan, G.; Gillet, W.; Zhang, S.

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, J.E.W.

A:Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.

A:Reference number: AB2577; PMID:11743193

A:Accession: AB2587

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-913 <KUR>

A:Cross-references: GB:AE008688; PIDN:AAL41112.1; PID:g17738404; GSPDB:GN00186

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: info

A:Map position: circular chromosome

C:Superfamily: translation initiation factor IF-2; translation elongation factor Tu homolog

Query Match 11.1%; Score 146; DB 2; Length 913;

Best Local Similarity 25.2%; Pred. No. 0.034;

Matches 65; Conservative 34; Mismatches 79; Indels 80; Gaps 12;

QY 29 PPALCARP-----VKEPRGL-----SAASPLLAETGAPRRFRSVPGRG-----AAGAV 72

DB 72 PAPAARPAQPPRIHQPGQOQRPGSSQSQRSQSSAPQOQRQADRPGRNVLDLSAGEM 131

QY 73 QELARALAHLEA-----ERQPRARAEQAEQOARVLQAQLLRVWGPNSDPALGLD 126

DB 132 EARRA---LMEAQARDVQAKQAEDEARKVEEQRIAAE--KMEANRAAEAEAAAK 186

QY 127 DDPDPAQAARALLRRLRDLAALQVLPAPMAALRPFPVDDGPAGPDAEEAGDE 186

DB 187 VAASQPAEV-----RAEPASERPAAPARTDAPQ-----SAAAPRSAPA--- 231

QY 187 TPVDPELLRYLLGRILAGSADSEGVAAAPRLRRADHD-----VGSELPPGVLGAL 239

DB 232 TPD-----AAAPRRRTGGDEDDRGAVRRGSSLPARG----- 264

QY 240 LRVKRLTETPAQVPARRL 257

DB 265 ----KVAPAPAKPAARL 278

RESULT 3

C75318

hypothetical protein - *Deinococcus radiodurans* (strain RL)

C:Species: *Deinococcus radiodurans*

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000

C:Accession: C75318

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* RL.

A:Reference number: A75250; MUID:20036896; PMID:10567266

A:Accession: C75318

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-553 <WHI>

A:Cross-references: GB:AE002044; GB:AE000513; NID:g6459872; PIDN:AAF11639.1; PID:g645988

A:Experimental source: strain RL

C:Genetics:

A:Gene: DR2090

A:Map position: 1

Query Match 10.2%; Score 134; DB 2; Length 553;

Best Local Similarity 25.0%; Pred. No. 0.12;

Matches 70; Conservative 24; Mismatches 108; Indels 78; Gaps 11;

QY 24 GLFRPPALCARPVKEPRG-----LSAASPP-----LAET 53

DB 32 GTLAPAPQIVRRVRRPAGOVPTTISALSGVPEPPAPRAEVOVRAVDRPSRRTPATAEEP 91

A:Accession: T33110
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-460 <TIN>
A:Cross-references: EMBL:AF067607; PIDN:AA017641.1; GSPDB:GN00022; CESP:C18H7.3
A:Experimental source: strain Bristol N2; clone C18H7
C:Genetics:
A:Gene: CESP:C18H7.3
A:Map position: 4
A:Introns: 84/1
C:Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Query Match 9.7%; Score 127.5; DB 2; Length 460;
Best Local Similarity 30.0%; Pred. No. 0.27;
Matches 64; Conservative 10; Mismatches 84; Indels 55; Gaps 8;
QY 29 PP-----ALCARPVKEPRGLSASPLAETGAPRRFRSVPGRGEAGVQELARALAHLEA 85
Db 275 PPSAALGAGGGAEPAG-----AAPEAAAAAPEAAPEAAAGAGGAGGAGPAGAAAPDAAA 330
QY 86 ERQERARAEQAEQDQARVLAQLLRVWGAPRNSDPALGLDD-----DPDPAQAQLARALL 141
Db 331 AAPEAAPAEAPAAE-----GAGGGAEPAGAAPDAAAAAPEAAPEAAAPAAE 377
QY 142 RA--RLDPAALAAQLVPAPVPAALRPRPPVYDDGAPGDAEEAGDETDPDPELLRYLL 199
Db 378 GAGGGAEPAGAAAPDAAAPEAAPE-----AAPAAEGAG----- 414
QY 200 GRILAGSADSEGVAAAPRLRRRAADHDVGSLEPP 232
Db 415 -----GGAEPAG-AAPEAAAAAPGAGGGERAPP 441

RESULT 6
B56708
extracellular signal-regulated kinase 5 - human
C:Species: Homo sapiens (man)
C:Date: 19-Oct-1995 #sequence_revision 19-Oct-1995 #text_change 24-Sep-1999
C:Accession: B56708
R;Zhou, G.; Bao, Z.Q.; Dixon, J.E.
J. Biol. Chem. 270, 12665-12669, 1995
A:Title: Components of a new human protein kinase signal transduction pathway.
A:Reference number: A56708; MUID:95279403; PMID:7759517
A:Accession: B56708
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-815 <ZHO>
A:Cross-references: GB:U25278; NID:g837260; PIDN:AAA81381.1; PID:g837261
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo
C:Keywords: ATP
F:52-315/Domain: protein kinase homology <KIN>
F:60-68/Region: protein kinase ATP-binding motif

Query Match 9.4%; Score 124.5; DB 2; Length 815;
Best Local Similarity 22.7%; Pred. No. 0.76;
Matches 67; Conservative 31; Mismatches 116; Indels 81; Gaps 11;
QY 5 PLWMPRAGGVGLLVLLGLFRPPA--LCARPVKEPRGLSASPLAETGAPRRFRS 62
Db 419 PSMWAPSGD-----CAMESPPAPPCGAPDTIDLTQPPPPVSEPPAPPKDKGA 469
QY 63 VPRGEAAGVQELARALAHLL-----EAERQERARAEQAEQD 100
Db 470 ISDNWTKAALKAAKLSRLSRDLRGPSAPLEAPRKPVTQAQRERERERERERERAKE 529
QY 101 QQAQLVLAQLLRVWGAPR-----NSDPALGL-----DDPDPAQAQLARALLRDLPAALAAQ 153
Db 530 REXRQRERKERKAGAGSGPSTDPALGLVLSND-----RSLERWTMRAR--PAAPALT 583
QY 154 LVPAPVPAALRPRPPVYDDGAPGDAEEAGDETDPDPELLRYLLGRILAGSADSEGA 213
Db 584 SVAPAPAPPTPTTTPVQPTSPPPGCPAQTGPQ-----PQSAGSTSGVPQPCPPPG-- 636

QY 214 APRRLRAADHDVGSLEPPGVLGALLRVKRLTAPQ-----VPARRLLPP 260
Db 637 -----PAPHTGPPGP-----IPVAPPQIATSTSLAAQSLVPP 671
RESULT 7
T29018
hypothetical protein ZK84.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
C:Accession: T29018
R;Kirsten, J.
submitted to the EMBL Data Library, April 1995
A:Description: The sequence of C. elegans cosmid ZK84.
A:Reference number: Z20553
A:Accession: T29018
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-801 <KIR>
A:Cross-references: EMBL:U23181; PIDN:AA048204.1; GSPDB:GN00020; CESP:ZK84.1
A:Experimental source: strain Bristol N2; clone ZK84
C:Genetics:
A:Gene: CESP:ZK84.1
A:Map position: 2
A:Introns: 22/2; 45/3; 108/1
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo

Query Match 9.3%; Score 123; DB 2; Length 801;
Best Local Similarity 28.5%; Pred. No. 0.93;
Matches 73; Conservative 14; Mismatches 107; Indels 62; Gaps 12;

QY 28 PPPALCARPVKEPRGLSASPLAETGAPRRFRSVPGRGEAA-----GAVQELARALAH 81
Db 507 PAPAPIEAPATDAATLETAPAAAE-----PAPAAEAAGYDAPSSVPEETPAPAP 557
QY 82 LLE-----AERQERARAEQAEQDQARVLAQLLRVWGAPRNSD---PALGLDDDP-DAP 132
Db 558 AADETPAPAAAEETPAPAAAEETPAPAAAEETPAPAAAEETPAPAAAEETPAPAAAEETPAPAP 617
QY 133 AAQLARALLRA-RLDPA--ALAAQLVPAPVPAALRPRP-----PVYDDGPA-GPDA 180
Db 618 AAETPAPAPAAVEETPAPAPAAVEETPAPAPAAVEETPAPAPAAVEETPAPAPAAVEETPAPAPAA 677
QY 181 EE-----AGDETDPDPELLRYLLGRILAGSADSEGVAAAPRRRLRAADHDVGSLEPP 234
Db 678 EETPAPAAAEETPAPAP-----AAETPAPAAAEETPAPAPAAAEETPAPAPAAAEAPAPA 724
QY 235 VLGALLRVKRLTAPAP 250
Db 725 PAAP-----ETPAP 733

RESULT 8
T03465
Probable exonuclease (EC 3.1.15.-) - Rhodobacter capsulatus
C:Species: Rhodobacter capsulatus
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 02-Jun-2000
C:Accession: T03465
R;Vicek, C.; Paces, V.; Maltsev, N.; Paces, J.; Haselkorn, R.; Fonstein, M.
Proc. Natl. Acad. Sci. U.S.A. 94, 9384-9388, 1997
A:Title: Sequence of a 189-kb segment of the chromosome of Rhodobacter capsulatus SBI
A:Reference number: 214955; MUID:97404404; PMID:9256491
A:Accession: T03465
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-1238 <VLC>
A:Cross-references: EMBL:AF010496; NID:g3128256; PIDN:AA016118.1; PID:g3128266
C:Genetics:

A:Gene: sbcc
A:Map position: 1
C:Superfamily: chromosome segregation protein SMC1
C:Keywords: DNA repair; exonuclease; hydrolase

QY	192	PELLRYLLGRILAGSADSEGVAPRRLRR---	AADHVGSELPEPGVLGALLVRKLETP	248
		: : : : : : : :		
Db	288	P-LVRKLAENGVDLTVRGTVGGVRIRKQDVAAAAA	AKAAAP-----APAA	334

QY	RSYPRGEAGAVQELARALAHLLLEAEQRERARAEQEA-EDQQRVLQALLRWVWGAPRNS	119
Db	PSAPRAAAAA--AARRSASSSSSSSSSSSSSSSEGEDEGVRGCAPIARAGPPP----	426

QY 120 DPALGLDDDPAPAAQLARALLRARLLDPAAALQOLVPAPVPAALRRPP----- 169
Db 427 -----SPPAPAAAPRPSASSASATSSAAAAPAPAPEFA-----RPPRRKRSTNNH 473
QY 170 --VYDDGPA---GDAEAGDETPDVDPDEL---LRYLLGRILAGSADS-EGV-----A 213
Db 474 LSLMADGPPPTDGLPLTGLPEWPGSDPPADGRVY-----GGAGDSRGLWDEDDVRQ 527
QY 214 APRRLRAA-----DHDVGSLEPPEGVGLGALLRVKRLTETAPQVPA 256
Db 528 AAARYRAAGPVPVFIPEMGDSRQKHEALVRLIYSGAAGEAM--SWLQNPROMQAPDQR 583

RESULT 12

EDBE75

immediate-early protein IE175 - human herpesvirus 1
C;Species: human herpesvirus 1
C;Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 16-Jun-2000
C;Accession: A23510
R;McGeoch, D.J.; Dolan, A.; Donald, S.; Brauer, D.H.K.
Nucleic Acids Res. 14, 1727-1745, 1986
A;Title: Complete DNA sequence of the short repeat region in the genome of herpes simplex virus 1
A;Reference number: A23510; MUID:86148504; PMID:3005980
A;Accession: A23510
A;Molecule type: DNA
A;Residues: 1-1298 <MCG>
A;Cross-references: GB:X14112; GB:D00317; GB:D00374; GB:S40593; NID:g1944536; PIDN:CAA32
C;Comment: This protein acts at the transcriptional regulatory level and is required for the establishment of latency.
C;Genetics:
A;Gene: IE3
A;Map position: short repeat region (IR-s)
A;Superfamily: herpesvirus immediate-early protein IE175
C;Keywords: DNA binding; early protein; transcription regulation

Query Match 9.0%; Score 118.5; DB 1; Length 1298;
Best Local Similarity 24.9%; Pred. No. 3;
Matches 88; Conservative 24; Mismatches 121; Indels 121; Gaps 15;

QY 2 AGSP-LWGPAGGVLVLLLLGLFRPPPPALCARPVKEP----- 41
Db 355 SGAPAAVWAPELGDAQOYALITRLTTPDAEAMGWLQNPVPGDVALDQACFRISGAA 414
QY 42 -----GLSASPPLAETGAPRRFRSVPCEAGAVQ-----ELARAL 79
Db 415 RNSSFTTGSVARVPHLYGMAAGREGWGLAHAAAVAMSRDYDRAQKGLTSLRAY 474
QY 80 AHLEAER-----QERARAEAEEDQ-----ARVLAQLLRVWGAP 116
Db 475 APLARENALTGAGSGAGADEGVAVAAMAPGERAVPAGYGAAGILALAGRLSAAP 534
QY 117 RNSDPALGLDDDPADP-----AAQLAR-----ALLRARLDPAAALQOLVPAPV 159
Db 535 --ASPAGG--DDPDAARHADADDAGRRAGQAGRVAVECLAACRCILALAEFGDGLAAV 590
QY 160 PA-AALRPPPVYDDGAPGDAEAGDETPDVDPDELRYLLGRI----- 202
Db 591 PGLAGARPAPPRPEGPAGP-----ASPPPHADAPRLRAWLRELFRVDRDALVLMRLRGDL 646
QY 203 -LAGSADSEGVAPRRRLRRADHDVGSLEPPEGVLG-ALLRVKRLTETAPQVPA 254
Db 647 RVAGGSEA-AVAARAVSLVA-----GALGPALPRDPRLPSSAAAAA 688

RESULT 13

A45344

immediate-early protein - suid herpesvirus 1 (strain Kaplan)
C;Species: suid herpesvirus 1
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Jul-1999
C;Accession: A45344
R;Vicek, C.; Kozmik, Z.; Paces, V.; Schirm, S.; Schwzyer, M.
Virology 179, 365-377, 1990
A;Title: Pseudorabies virus immediate-early gene overlaps with an oppositely oriented op

A;Reference number: A45344; MUID:91021039; PMID:2171211
A;Accession: A45344
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-1446 <VLC>
A;Cross-references: GB:M34651; NID:g334070; PIDN:AAA47470.1; PID:g334071
C;Superfamily: herpesvirus immediate-early protein IE175
C;Keywords: DNA binding; early protein; transcription regulation

Query Match 8.9%; Score 118; DB 1; Length 1446;
Best Local Similarity 26.6%; Pred. No. 3.6;
Matches 83; Conservative 22; Mismatches 97; Indels 110; Gaps 15;

QY 16 GLLVLLLLGLFRPPPPALCARPVKEPGLSASPPPLAE-TGAPRRFRSVP-----RGEA 68
Db 696 GVLERLL-----PCPLRLPAPARAPAAALGPAC--LEEVTAALLRALDAIPGAPERRQA 748
QY 69 AGAVOELARALAHLEAERQERARAEAEEDQOARVLAQLLRVWGAPRNSDPALGLDD 128
Db 749 ADSVALVARTVAPLV---RYSVDGARAREAAWTAAALF-----APAN----- 788
QY 129 PDAPAAQALARALLRARLDPAAALAAQL-----VPAPVPAALRRPPPVYDDGAPG 177
Db 789 --VAARLAEA--AARCPAPAPCLPLMPWEPQCLVVPAPAPAAAGAP-SGLPCSGPSS 843
QY 178 PDAAEAGDET-----PDVDELLRY----- 197
Db 844 PASTKSSSTKSSSTKSSSTKSSGSGYASSPAAGDPAPAPERKKRRAPGARRPGDEDE 903
QY 198 -LLGRILAGSA---DSEGVAAAPRLRAADHDVGSLEPPEGVLGALL-----RV 242
Db 904 GLSAAALRGDGHGRDDEEDRGPRKRRSLGLGPAPDPAP-----ALLSSSSSSSEDDRL 958
QY 243 KRLETPAPQVPA 254
Db 959 RRLPGMPHEPA 970

RESULT 14

FOLJLK

gag polyprotein - simian foamy virus (type 3, strain LK3)
N;Alternate names: core polyprotein
C;Species: simian foamy virus
A;Note: host (African green monkey)
C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 08-Apr-1994
C;Accession: A40820
R;Renne, R.; Friedl, E.; Schweizer, M.; Fleps, U.; Turek, R.; Neumann-Haefelin, D.
Virology 186, 597-608, 1992
A;Title: Genomic organization and expression of simian foamy virus type 3 (SFV-3).
A;Reference number: A40820; MUID:92124734; PMID:1310187

A;Accession: A40820
A;Molecule type: DNA
A;Residues: 1-643 <REN>
A;Cross-references: GB:M74895
C;Genetics:
A;Gene: gag
C;Superfamily: foamy virus gag polyprotein
C;Keywords: core protein; polyprotein

Query Match 8.9%; Score 117; DB 1; Length 643;
Best Local Similarity 22.2%; Pred. No. 1.8;
Matches 66; Conservative 40; Mismatches 101; Indels 90; Gaps 14;

QY 8 WGPAGGVGLVLLLLGLFRPPPPALCARPVKEPGLSAAAP--PLAETGAPRRFRSVP 65
Db 39 WGPCTR-----YLLVSIFLQDDSS--CQPLQOPWRPREGRPVNPVINTI-----BAPW 84
QY 66 GERAGAVOELARALAHLEAERQERARAEAEEDQOARVLAQLLRVWGAPRNSDPALGLDD 102
Db 85 GERLQAFEDLDVABGTLRFGPLANGWIPGDEYSMEFPPLAQIAQMQORDELEIIDLIT 144
QY 103 ARVLAQLL-----RWGAPRNSDDPALGLDD-----PDAPAAQALARALLRARL 145

Db 145 GOICAQVIDVMDQDAIRGLERRIQDLRLGRLNLPVAGIQAPPPSPICQ----- 194
QY 146 DPAALAAQLVPAPVPAALRPFPVYDDGAPGPDAAEEAGDETP-DVDPPELLRYLLGRILA 204
Db 195 -----PIASSSLQIP-----GSSSPADLDGIWTPRQIDPRLSRVAYNPFLP 237
QY 205 GSADSEGVAAPRRLRRAADHDVGSELP--PEGVLGALLR-VKLETPAQ-VPARRL 257
Db 238 GSSDGGSGSIP--VQSPAPVLPVLPSPVQFVPIQYVAQPPVPAPQAIPIQHI 292

RESULT 15
TI4267
Xin protein, stage early embryo - mouse
C:Species: Mus musculus (house mouse)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: TI4267
R:Wang, D.Z.; Lin, J.J.C.
submitted to the EMBL Data Library, March 1998
A:Description: Involvement of a novel gene, Xin, in cardiac looping.
A:Reference number: Z17948
A:Accession: TI4267
A>Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-1677 <WAN>
A:Cross-references: EMBL:AF051945; NID:g2970645; PID:g2970646; PIDN:AAC06023.1
A:Experimental source: cardiac muscle; stage early embryo

Query Match 8.8%; Score 116.5; DB 2; Length 1677;
Best Local Similarity 25.9%; Pred No. 5.2;
Matches 63; Conservative 26; Mismatches 113; Indels 41; Gaps 10;

QY 28 PPPALCARPVKEPRLSAAASPPLAETGAPRRFRSRVPRGEAAGAVQE-----LARALAH 81
Db 993 PPPSRAAGKVLPLNGKPVAAQLQEARKKRDISHAGQKKAASGRPEGTIASPIGSGAPD 1052
QY 82 LLEAEQER-ARAEQAEEDQARVLAQLLRVWGAPRNSDP-ALGLDDDDPDAPAAQLARA 139
Db 1053 LQEAQNLRLATAEAGSLHQQ---VLSR-----HPQSGDPVATSMVPQDVQLQASTPATG 1103
QY 140 LLRLRLDPAALAAQLVPAPVPAALRPFPVYDDGAPGPDAAEEAGDETPDVDPPELLRYLL 199
Db 1104 VTQGSISAVAGSEARIPAVPQKAATVDP---DHTQGHQEDSIOQAPEPLQEPFLLIH 1160
QY 200 GRILAGSADSEGVAAAPRRLRRAADHDVGSELPPEGVLGALLRVKRLKETPAPQVPAR--RL 257
Db 1161 NR-----PSGQKTP-----GSETKPSKAESTML--PRKKPPVPKPAHLSQI 1201
QY 258 LPP 260
Db 1202 HPP 1204

Search completed: April 4, 2003, 14:18:38
Job time : 22 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2003, 14:09:29 ; Search time 11 Seconds
(without alignments)
980.350 Million cell updates/sec

Title: US-09-803-589-6

Perfect score: 1319

Sequence: 1 MAGSPLLWCPAGGVGLLV.....RVKRLTAPAPQVPARLLPP 260

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	124.5	9.4	815	1 MK07_HUMAN	Q13164 homo sapien
2	123	9.3	1238	1 SBCC_RHOCA	O68032 rhodobacter
3	119.5	9.1	1461	1 IE18_PRVIF	P11675 pseudobacter
4	118.5	9.0	1298	1 ICPA_HSV11	P08392 herpes simp
5	118	8.9	1446	1 IE18_PRVKA	P33479 pseudorabie
6	117	8.9	643	1 GAG_SFV3L	P27400 simian foam
7	116.5	8.8	616	1 VGF_HUMAN	O15240 homo sapien
8	116	8.8	548	1 ERF_HUMAN	P50548 homo sapien
9	116	8.8	551	1 ERF_MOUSE	P70459 mus musculus
10	113.5	8.6	806	1 MK07_MOUSE	Q9WV88 mus musculus
11	113.5	8.6	933	1 PRGR_HUMAN	P06401 homo sapien
12	113.5	8.6	3164	1 TEGU_HSV11	P10220 herpes simp
13	112	8.5	817	1 NEB2_RAT	O35274 rattus norv
14	111	8.4	2142	1 BAT2_HUMAN	P48634 homo sapien
15	110	8.3	1411	1 TCOF_HUMAN	Q13428 homo sapien
16	109.5	8.3	2564	1 SPCQ_HUMAN	Q9H254 homo sapien
17	108	8.2	319	1 FXE3_HUMAN	Q13461 homo sapien
18	107	8.1	676	1 ICPO_HSVBK	P29836 bovine herp
19	107	8.1	954	1 M3KA_HUMAN	Q02779 homo sapien
20	107	8.1	2517	1 NCR2_HUMAN	Q9Y618 h nuclear r
21	106.5	8.1	395	1 TI22_HUMAN	Q9Y3q8 homo sapien
22	106	8.0	703	1 ML51_HUMAN	O15234 homo sapien
23	105	8.0	622	1 3BP1_HUMAN	Q9Y313 homo sapien
24	105	8.0	827	1 MOF_DROME	O02193 drosophila
25	105	8.0	1532	1 IGA_NEIGO	P09790 neisseria q
26	103.5	7.8	336	1 YB56_XENLA	P21574 xenopus lae
27	103.5	7.8	439	1 XP2_XENLA	P17437 xenopus lae
28	103.5	7.8	1184	1 PC12_HUMAN	Q9N994 homo sapien
29	103	7.8	603	1 NMBL_MOUSE	O08919 mus musculus
30	103	7.8	3301	1 CLR3_MOUSE	Q91210 mus musculus
31	102.5	7.8	553	1 ODO2_MYCTU	O10381 mycobacteri
32	102.5	7.8	628	1 V70K_TYMYA	P20131 turnip yell
33	102.5	7.8	1402	1 IF4G_RABIT	P41110 oryctolagus

ALIGNMENTS

RESULT 1

ID	MK07_HUMAN	STANDARD;	PRT;	815 AA.
AC	Q13164; Q16634;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Mitogen-activated protein kinase 7 (EC 2.7.1.-) (Extracellular signal-			
DE	regulated kinase 5) (ERK-5) (ERK4) (BMK1 Kinase).			
GN	MAPK7 OR PRK47 OR ERK5 OR ERK4.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Fetal brain;			
RX	MEDLINE=95279403; PubMed=7759517;			
RA	Zhou G., Bao Z.Q., Dixon J.E.;			
RT	"Components of a new human protein kinase signal transduction			
RT	pathway.";			
RL	J. Biol. Chem. 270:12665-12669(1995).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Placenta;			
RX	MEDLINE=95374539; PubMed=7646528;			
RA	Lee J.-D., Ulevitch R.J., Han J.;			
RT	"Primary structure of BMK1: a new mammalian map kinase.";			
RL	Biochem. Biophys. Res. Commun. 213:715-724(1995).			
CC	-1- FUNCTION: MEK5 AND ERK5 INTERACT SPECIFICALLY WITH ONE ANOTHER AND			
CC	NOT WITH MEK1/ERK1 OR MEK2/ERK2 PATHWAYS.			
CC	-1- ENZYME REGULATION: ACTIVATED BY TYROSINE AND THREONINE			
CC	PHOSPHORYLATION (BY SIMILARITY).			
CC	-1- TISSUE SPECIFICITY: EXPRESSED IN MANY ADULT TISSUES. ABUNDANT IN			
CC	HEART, PLACENTA, LUNG, KIDNEY AND SKELETAL MUSCLE. NOT DETECTABLE			
CC	IN LIVER.			
CC	-1- DOMAIN: THE SECOND PROLINE-RICH REGION MAY INTERACT WITH ACTIN			
CC	TARGETING THE KINASE TO A SPECIFIC LOCATION IN THE CELL.			
CC	-1- PTM: AUTOPHOSPHORYLATED ON THREONINE AND TYROSINE RESIDUES, WHEN			
CC	THE C-TERMINAL PART OF THE KINASE, WHICH COULD HAVE A REGULATORY			
CC	ROLE, IS ABSENT.			
CC	-1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.			
CC	MAP KINASE SUBFAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; U25278; AAA81381.1; -			
DR	EMBL; U29725; AAA82931.1; -			
DR	EMBL; U29726; AAA82932.1; -			
DR	EMBL; U29727; AAA82933.1; -			

Q9ukn7 homo sapien
P28284 herpes simp
Q9jlvl1 mus musculus
Q04637 homo sapien
P26651 homo sapien
P13627 paracoccus
P03972 bos taurus
Q96rd6 homo sapien
P29128 bovine herp
Q05152 oryctolagus
P74876 salmonella
Q9c0c2 homo sapien

34 102.5 7.8 3530 1 MY15_HUMAN
35 102 7.7 825 1 ICPO_HSV2H
36 101.5 7.7 577 1 BAG3_MOUSE
37 101.5 7.7 1395 1 IF4G_HUMAN
38 101 7.7 326 1 TTP_HUMAN
39 101 7.7 450 1 CYL_PANDE
40 101 7.7 575 1 MIS_BOVIN
41 101 7.7 633 1 PAN2_HUMAN
42 101 7.7 676 1 ICPO_HSVBJ
43 101 7.7 2339 1 CCAB_RABIT
44 100.5 7.6 642 1 DP3X_SALTY
45 100 7.6 1729 1 TABP_HUMAN

OC Viruses: dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=31523;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=89315207; PubMed=2546124;
RA Cheung A.K.;
RT "DNA nucleotide sequence analysis of the immediate-early gene of
RT pseudorabies virus.";
RL Nucleic Acids Res. 17:4637-4646(1989).
RN [2]
RN REVISIONS.
RN Cheung A.K.;
RA Submitted (NOV-1989) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THIS IE PROTEIN IS A MULTIFUNCTIONAL PROTEIN CAPABLE
CC OF MIGRATING TO THE NUCLEUS, BINDING TO DNA, TRANS-ACTIVATING
CC OTHER VIRAL GENES, AND AUTOREGULATING ITS OWN SYNTHESIS.
CC -1- SUBCELLULAR LOCATION: NUCLEUS OF INFECTED CELLS.
CC -1- PTM: A LONG STRETCH OF SERINE RESIDUES MAY BE A MAJOR SITE OF
CC PHOSPHORYLATION.
CC -1- SIMILARITY: BELONGS TO THE ICP4/IE140/IE180 FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X15120; CAA33214.1;
DR PIR: S04713; EDBE7F.
DR InterPro: IPR005205; Herpes_ICP4_C.
DR InterPro: IPR005206; Herpes_ICP4_N.
DR Pfam: PF03584; Herpes_ICP4_N.
DR Pfam: PF03585; Herpes_ICP4_C.
KW Early protein; Transcription regulation; Trans-acting factor;
KW DNA-binding; Phosphorylation; Nuclear protein.
FT DOMAIN 390 405 POLY-SER.
FT DOMAIN 958 966 POLY-SER.
SQ SEQUENCE 1461 AA; 149833 MW; 7F31E7ABE403B208 CRC64;

Query Match 9.1%; Score 119.5; DB 1; Length 1461;
Best Local Similarity 25.1%; Pred. No. 1.9;
Matches 77; Conservative 27; Mismatches 105; Indels 98; Gaps 13;
QY 16 GLLVLLLLGLFRPPPPALCARPVKEPRGLSASPPPLAE--TGAPRRFRRSVP-----RGEA 68
DB 707 GVLERLL-----PCPLRLPAPAPAPAAALGPAC--LEEVTAALLALRAIPGAGPAERQQA 759
QY 69 AGAVQELARALAHLEERQERARAEAOEADQOARVL-----AOLLRVWGAPRNSD 120
DB 760 ADSVALVARTVAPLV---RYSVDGARAREAAWTYAAALFAPANVAGARLAEEAARPGPAE 816
QY 121 PALGLDD--DPDAPAAQLARALLRARLDPAALAAQLVPAPVPAALRPRPPVYDDGPAGPD 179
DB 817 PAPGLPPLWPEQGL-----VVPAPAPAAAGAP--SGLPLSGSPSPA 856
QY 180 AEEAGDET-----PDVDELLRY-----L 198
DB 857 STKSGSTKSSGKSGLSGSSGYASSPAAGPDPAPEPRKKRRAPGARRRPGDEEGL 916
QY 199 LGRILASA-----DSEGAAPRRRLRAADHDVGSLELPEGVIGALL-----RVKRL 247
DB 917 SGSALRGDGHGRDDEEDRGPRRRRRSLGLGPADPDAPALVSSSSSSSSSSDRLRLRPLG 976
QY 248 PAPQVPA 254
DB 977 PMPEHPA 983
RESULT 4
ICP4_HSV11

ICP4_HSV11 STANDARD; PRT; 1298 AA.
P08392;
01-AUG-1988 (Rel. 08, Created)
01-AUG-1988 (Rel. 08, Last sequence update)
15-JUL-1998 (Rel. 36, Last annotation update)
Trans-acting transcriptional protein ICP4 (Transcriptional activator
IE175) (Alpha-4 protein).
GN ICP4 OR IE175 OR RSL.
OS Herpes simplex virus (type 1 / strain 17).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10299;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=88274327; PubMed=2839594;
RA McGeoch D.J., Dalrymple M.A., Davison A.J., Dolan A., Frame M.C.,
RA McNab D., Perry L.J., Scott J.E., Taylor P.;
RT "The complete DNA sequence of the long unique region in the genome of
RT herpes simplex virus type 1.";
RL J. Gen. Virol. 69:1531-1574(1988).
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE=86148504; PubMed=3005980;
RA McGeoch D.J., Dolan A., Donald S., Brauer D.H.K.;
RT "Complete DNA sequence of the short repeat region in the genome of
RT herpes simplex virus type 1.";
RL Nucleic Acids Res. 14:1727-1745(1986).
RN [3]
RN DNA-BINDING DOMAIN.
RX MEDLINE=90174974; PubMed=2155403;
RA Wu C.-L., Wilcox K.W.;
RT "Codons 262 to 490 from the herpes simplex virus ICP4 gene are
RT sufficient to encode a sequence-specific DNA binding protein.";
RL Nucleic Acids Res. 18:531-538(1990).
RN [4]
RN INFLUENCE OF PHOSPHORYLATION ON FUNCTION.
RX MEDLINE=9112047; PubMed=1846804;
RA Papavassiliou A.G., Wilcox K.W., Silverstein S.J.;
RT "The interaction of ICP4 with cell-infected-cell factors and its
RT state of phosphorylation modulate differential recognition of leader
RT sequences in herpes simplex virus DNA.";
RL EMBO J. 10:397-406(1991).
CC -1- FUNCTION: THIS IE PROTEIN IS A MULTIFUNCTIONAL PROTEIN CAPABLE
CC OF MIGRATING TO THE NUCLEUS, BINDING TO DNA, TRANS-ACTIVATING
CC OTHER VIRAL GENES, AND AUTOREGULATING ITS OWN SYNTHESIS. BINDING
CC OF ICP4 TO ALPHA GENES CAUSES REPRESSION OF THEIR EXPRESSION.
CC ICP4 ALSO ACTS AS A TRANSCRIPTIONAL ACTIVATOR OF BETA AND GAMMA
CC GENES. ICP4 BINDS WITH HIGH AFFINITY TO THE SEQUENCE 5'-ATCGTC-3'.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: NUCLEUS OF INFECTED CELLS.
CC -1- PTM: ICP4 IS ADP-RIBOSYLATED.
CC -1- PTM: THE LONG STRETCH OF SER IS A MAJOR SITE OF PHOSPHORYLATION.
CC ONLY THE PHOSPHORYLATED FORMS OF ICP4 ARE CAPABLE OF INTERACTING
CC WITH BETA OR GAMMA GENES.
CC -1- SIMILARITY: BELONGS TO THE ICP4/IE140/IE180 FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L00036; AAA96675.1;
DR EMBL: L00036; AAA96688.1;
DR EMBL: X14112; CAA32286.1;
DR EMBL: X14112; CAA32278.1;
DR EMBL: X06461; CAA29763.1;
DR PIR: A23510; EDBE75.
DR InterPro: IPR005205; Herpes_ICP4_C.
DR InterPro: IPR005206; Herpes_ICP4_N.
DR Pfam: PF03584; Herpes_ICP4_N.

```
DR Pfam: PF03585; Herpes_ICP4_C; 1.
KW Early protein; Transcription regulation; Trans-acting factor;
KW DNA-binding; Phosphorylation; Nuclear protein; ADP-ribosylation.
FT DOMAIN 176 199
FT DNA_BIND 262 490
FT SEQUENCE 1298 AA; 132843 MW; 4F32E04C95CA9344 CRC64;
SQ
Query Match 9.0%; Score 118.5; DB 1; Length 1298;
Best Local Similarity 24.9%; Pred. No. 1.9;
Matches 88; Conservative 24; Mismatches 121; Indels 121; Gaps 15;
QY 2 AGSP-LIWMGPRAGGVLLVLLGLFRPPPPALCARPVKEPR----- 41
DB 355 SCAPAAVWPELDGAAQVVALITRLTYTPDAEMGWLNQPRVVGVDALDQACFRISGAA 414
QY 42 -----GLSAASPPLAETGAPRRFRSVPGRGEAGAVQ-----ELARAL 79
DB 415 RNSSSPITGSVARVPHGLGYAMAAGFCGWLAAHAAVAMSRDYRAQKGFLLTSLRAY 474
QY 80 AHLLEAER-----QERARAEQAEDEQ-----ARVLAQLLRVWVGAP 116
DB 475 APLLARENAALTGAAGSPGAGADDEGVAVAATAAAGERAVPAGYGAAGTILAAALGRLSAAP 534
QY 117 RNSDPALGDDDDPADP-----AAQLAR-----ALLRLRDPALAAQLVPAPV 159
DB 535 --ASPAGG--DDPDRAHRAHADDDAGRRRAQAGRVAVECUACRGILAEFGDGLAAV 590
QY 160 PA-AALRPVPPVDDGPAGPADEAGDETFDVPDELLRYLLGRI----- 202
DB 591 PGLAGARPASPPRPEGPAGP-----ASPPPHADAPRLRAWLRELFVRDVALVLMRLRGDL 646
QY 203 -LAGSADSEGVAAAPRLRRAADHDVGSSELPPEGVIG-ALLRVKRLTETAPQVPA 254
DB 647 RVAGGSEA-AVAARAVSLVA-----GALGPALPRDPRPLPSSAAAAA 688
RESULT 5
IE18_PVKVA
ID IE18_PVKVA STANDARD; PRT: 1446 AA.
AC P33479;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Immediate-early protein IE180.
GN IE.
OS Pseudorabies virus (strain Kaplan) (PRV).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=33703;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91021039; PubMed=2171211;
RA Vilek C., Kozmik Z., Paces V., Schirm S., Schwytzer M.;
RT "Pseudorabies virus immediate-early gene overlaps with an oppositely
RT oriented open reading frame: characterization of their promoter and
RT enhancer regions.";
RL Virology 179:365-377(1990).
CC -!- FUNCTION: THIS IE PROTEIN IS A MULTIFUNCTIONAL PROTEIN CAPABLE
CC OF MIGRATING TO THE NUCLEUS, BINDING TO DNA, TRANS-ACTIVATING
CC OTHER VIRAL GENES, AND AUTOREGULATING ITS OWN SYNTHESIS.
CC -!- SUBCELLULAR LOCATION: NUCLEUS OF INFECTED CELLS.
CC -!- PTM: A LONG STRETCH OF SERINE RESIDUES MAY BE A MAJOR SITE OF
CC PHOSPHORYLATION.
CC -!- SIMILARITY: BELONGS TO THE ICP4/IE140/IE180 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
DR EMBL: M34651; AAA47470.1; -.
DR PIR: A45344; A45344.
DR InterPro: IPR005205; Herpes_ICP4_C.
DR InterPro: IPR005206; Herpes_ICP4_N.
DR Pfam: PF03584; Herpes_ICP4_N; 1.
DR Pfam: PF03585; Herpes_ICP4_C; 1.
KW Early protein; Transcription regulation; Trans-acting factor;
KW DNA-binding; Phosphorylation; Nuclear protein.
FT DOMAIN 347 354
FT DOMAIN 379 397
FT POLY-SER.
SQ SEQUENCE 1446 AA; 148640 MW; 81F43A3DE3DDA068 CRC64;
Query Match 8.9%; Score 118; DB 1; Length 1446;
Best Local Similarity 26.6%; Pred. No. 2.3;
Matches 83; Conservative 22; Mismatches 97; Indels 110; Gaps 15;
QY 16 GLIVLLLLGLFRPPPPALCARPVKEPRGLSAASPPLAE--TCAPRRFRSVP-----RGEA 68
DB 696 GVLRL-----PCRLPAPAPAPAAALGPAC--LEEVTAALLALRDAIPGAGPAERROA 748
QY 69 AGAVOELARALAHLEAEQERARAEQAEDEQARVLAQLLRVWVGAPRNSDPALGLDDD 128
DB 749 ADSVALVARTVAPLV--RYSVDGARAREANTYAAALF-----APAN----- 788
QY 129 PDAPRAQAARALLRLRDLPAALAAQL-----VPAPVPAALRRPPVYDDGPAG 177
DB 789 --VAARLAEA--AARPGPAEPAGLPPLWPEQGLVVPAPAPAAAGAP-SGLPGSGSS 843
QY 178 PDAAEAGDET-----POVDPELLRY----- 197
DB 844 PASTKSSSTKSSSTKSSGSSGYASSPAAGPDPAPERRKKRRAPGARRPGDEE 903
QY 198 -LLGRILAGSA-----DSEGVAAPRLRRAADHDVGSSELPPEGVIGALL-----RV 242
DB 904 GLSGAALRGDGHGRDDEEDGRKRRSLGLGPADPPAP-----ALLSSSSSSSEDDRL 958
QY 243 KRLETPAPQVPA 254
DB 959 RRLPGMPPEHPA 970
RESULT 6
GAG_SFV3L
ID GAG_SFV3L STANDARD; PRT: 643 AA.
AC P27400;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE GAG polyprotein (Core polyprotein).
GN GAG.
OS Simian foamy virus (type 3 / strain LK3) (SFV-3).
OC Viruses; Retroid viruses; Retroviridae; Spumavirus.
OX NCBI_TaxID=11644;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9214734; PubMed=1310187;
RA Renne R., Friedl E., Schweizer M., Fleps U., Turek R.,
RA Neumann-Haefelin D.;
RT "Genomic organization and expression of simian foamy virus type 3
RT (SFV-3).";
RL Virology 186:597-608(1992).
CC -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
DR EMBL: M74895; AAA47795.1; ALT_INIT.
DR PIR: A40820; FOLJLK.
```



```
DR InterPro: IPR004957; Spuma_gag.
DR Pfam: PF03276; Gag_spuma; 1.
KW Core protein; Polyprotein.
SQ SEQUENCE 643 AA; 69785 MW; C53A0575BA9B5949 CRC64;

Query Match      8.9%; Score 117; DB 1; Length 643;
Best Local Similarity 22.2%; Pred. No. 1.2; Indels 90; Gaps 14;
Matches 66; Conservative 40; Mismatches 101;

QY 8 WGRAGVGVLVLLGLLGRPPPPALCARPVKEPRGLSAA--PLAETGAPRRFRSVPVR 65
DB 39 WGPCTR-----YILVSIFLQDSS--GQLQPPWRPEGRVNPVHNTI-----EAPW 84
QY 66 AGAGAVOELARALAH-----LEAERERARAEQAEDQ 102
DB 85 GELQAFEDLDVAEGTLRFGLANGWIPGDEYSMEFQPPPLAQEIAQMORDEEILDI 144
QY 103 ARVLAQL-----RVMGAPRNSDPALGLDD-----PDAPAAQLARALLRRL 145
DB 145 GQICAQVIDLVMDQAIQIRLERIQRGLRNLVPVAGIAPPSPIGO----- 194
QY 146 DPAALAAQLVPAPVPAALPRPPVYDDGPAQDAEAGDETP-DVDPELLRYLLGLRLA 204
DB 195 -----PIASSSLQIP-----GSSSPADLDLGIWTQIDPRLSRVAYNPF 237
QY 205 GSADSEGVAPRLRRRAADHDVSELP--PEGVLGALLR-VKRLTPAPO-VPARLL 257
DB 238 GSSDGGSGSIP--VQPSAPPAVLPSLPSPVQSQTIOYVQAPVPPVAPQAIPQI 292

RESULT 7
VGF_HUMAN
ID VGF_HUMAN STANDARD; PRT; 616 AA.
AC O15240;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Neurosecretory protein VGF precursor.
GN VGF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=98008940; PubMed=9344675;
RA Canu N., Possenti R., Ricco A.S., Rocchi M., Levi A.;
RT "Cloning, structural organization analysis and chromosomal assignment
of the human gene for neurosecretory protein VGF.";
RL Genomics 45:443-446(1997).
CC -!- FUNCTION: MAY BE INVOLVED IN THE REGULATION OF CELL-CELL
CC INTERACTIONS OR IN SYNAPTOGENESIS DURING THE MATURATION OF THE
CC NERVOUS SYSTEM (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Stored in secretory vesicles and then
CC secreted (By similarity).
CC -!- TISSUE SPECIFICITY: Expressed in brain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Y12661; CAAT3210.1; -.
DR Genew; HGNC:12684; VGF.
DR MIM; 602186; -.
KW Growth factor; Signal.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 616 NEUROSECRETORY PROTEIN VGF.
FT DOMAIN 353 447 ASP/GLU-RICH (ACIDIC).
```

```
SQ SEQUENCE 616 AA; 67286 MW; CD1920610201BBB9 CRC64;

Query Match      8.8%; Score 116.5; DB 1; Length 616;
Best Local Similarity 25.9%; Pred. No. 1.2;
Matches 69; Conservative 27; Mismatches 109; Indels 61; Gaps 13;

QY 11 RAGVGVLVLLLL-GLFRPPPPALCARPVKEPRGLSAA-SPLAETGAPRRFRSVPVRGEA 68
DB 5 RLSASALFCILLGLGAAPP--GRPEAQPPPLSSHEKPEFVAGDAVP-----GPK 52
QY 69 AGAVOELARALAHLEAERERARAEQAEDQQAARVLAQ-LLRVMGAPRNSDPALGLDD 127
DB 53 DGAPEVRGA-----RNSEFQDEGELFGVDPRALAAVLLQALDRPASPAPSSQ 104
QY 128 DPAPAAQLARALLRRLRDLPAALAAQLVPAPVPAALPRPPVYDDGPAQDAEAGDET 187
DB 105 GPEERAAE--ALLTETVRSQTHSLPAAGEPEPAAPRPQTP--ENGP-----EASDPS 153
QY 188 PDVDPPELLRYLLGRILAGSADSGVAAPRLRRRAADHDVSELPPEGVIGALLRVKRL 247
DB 154 EELE-----ALASLLQELRDFSPSSAKRQETA-----AETETRTHTLTV-NLES 199
QY 248 PAPO-----VPARRLLPP 260
DB 200 PGPVWRASWGEFQARVPERAPLPP 225

RESULT 8
ERF_HUMAN
ID ERF_HUMAN STANDARD; PRT; 548 AA.
AC P50548; Q9UP17;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ETS-domain transcription factor ERF (Ets2 repressor factor).
GN ERF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96030784; PubMed=7598608;
RA Sgouras D.N., Athanasios M.A., Beal G.J. Jr., Fisher R.J., Blair D.G.,
RA Mavrothalassitis G.J.;
RT "ERF: an ETS domain protein with strong transcriptional repressor
activity, can suppress ets-associated tumorigenesis and is regulated
by phosphorylation during cell cycle and mitogenic stimulation.";
RL EMBO J. 14:4781-4793(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA Lamerdin J.E., McGready P.M., Skowronski E., Viswanathan V.,
RA Burkhardt-Schultz K., Gordon L., Dias J., Ramirez M., Stilwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Ganes J.,
RA Dangnan L., Erier A., Christensen M., Georgescu A., Avila J., Liu S.,
RA Attix C., Andreise T., Frankhelm M., Amico-Keller G., Coefield J.,
RA Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Krommiller B.,
RA Arellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A.,
RA Olsen A.S., Carrano A.V.;
RT "Sequence analysis of a 3.2 Mb region in 19q13.2 between CYP2F1 and
DL9S178.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: POTENT TRANSCRIPTIONAL REPRESSOR THAT BINDS TO THE H1
CC ELEMENT OF THE ETS2 PROMOTER. MAY REGULATE OTHER GENES INVOLVED
CC IN CELLULAR PROLIFERATION.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- PTM: PHOSPHORYLATED BY MULTIPLE KINASES INCLUDING PROBABLY ERK2.
CC PHOSPHORYLATION REGULATES THE ACTIVITY OF ERF.
CC -!- SIMILARITY: BELONGS TO THE ETS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```



```

QY      62 SVPRGEAGAVQELARAL--AHLLA-----ERQER 90
Db      468 -----EGAISDNTKAALKAAALLKSLRSLRDGPSAPLEAPRKVPVTAQREREKEE 520
QY      91 ARAEAQEADQOARVLQAQLLRVWGAPR-----NSDPALGL---DDPDPAQAQLARALL-- 141
Db      521 RRRQERAKERKRQERKERGKRGAGTGLGGPSTDPLAGLVLSND-----RSLLER 571
QY     142 --RARLDPAALAQAQLVAPVPAALARPVRPVVDDCGPAGDAEEAGDETPDVDPPELLRVLL 199
Db     572 WTRWARRPAP-----APAPAPAPAPASQAQTSTPTGPVSQSGTGLOP----- 615
QY     200 GRILAGSAD---SEGVAAPRLRRAADHDVGSELPEGVGALLRVRKLETPAQ----VP 253
Db     616 ----AGSIFGPASQPVCPP-----PGVPQPGAPIPAPLQT-----APSTSLLA 655
QY     254 ARBLPPP 260
Db     656 SQSLVPP 662

RESULT 11
PRGR_HUMAN STANDARD; PRT; 933 AA.
AC P06401; Q9UPF7;
CD 01-JAN-1988 (Rel. 06, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Progesterone receptor (PR).
GN PR OR NR3C3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RP MEDLINE=90228361; PubMed=2328727;
RA Kastner P., Krust A., Turcotte B., Stropp U., Tora L., Gronemeyer H.,
RA Chambon P.;
RT "Two distinct estrogen-regulated promoters generate transcripts
RT encoding the two functionally different human progesterone receptor
RT forms A and B." ;
RN EMBO J. 9:1603-1614(1990).
RN [2]
SEQUENCE FROM N.A.
RP MEDLINE=87184565; PubMed=3551956;
RA Misrahi M., Alger M., D'Auriol L., Loosfelt H., Meriel C.,
RA Fridlansky F., Guiochon-Mantel A., Galibert F., Milgrom E.;
RT "Complete amino acid sequence of the human progesterone receptor
RT deduced from cloned cDNA.";
RN Biochem. Biophys. Res. Commun. 143:740-748(1987);
RN [3]
SEQUENCE FROM N.A.
RP Krieback D.G., Agoulnik I.U., Tong X.-W.;
RN Submitted (JUL-1997) to the EMBL/GenBank/DBDJ databases.
RN [4]
X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 682-933.
RP MEDLINE=98282128; PubMed=9620806;
RA Williams S.P., Sigler P.B.;
RT "Atomic structure of progesterone complexed with its receptor.*";
RN Nature 393:392-396(1998).
CC -1 FUNCTION: THE REGULATION OF STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN IN
CC PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
CC -1 SUBCELLULAR LOCATION: Nuclear.
CC -1 ALTERNATIVE PRODUCTS: 2 ISOFORMS: A AND B (SHOWN HERE); ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1 DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC -1 SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC NR3 SUBFAMILY.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collabora
```


Handwritten scribble or mark.

Vertical handwritten marks on the right edge.

Vertical handwritten marks at the bottom right corner.

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2003, 14:15:40 ; Search time 15 Seconds
(without alignments)
509.997 Million cell updates/sec

Title: US-09-803-589-6
Perfect score: 1319
Sequence: 1 MAGSLLNGPRAGGVGLLV.....RVKRLTEPAQVPARRLLPP 260

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2.6/prodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2.6/prodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2.6/prodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2.6/prodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2.6/prodata/1/iaa/PCTUS_COMB.pep.*
- 6: /cgn2.6/prodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	137.5	10.4	882	4	US-09-413-814-78
2	119.5	9.1	248	3	US-08-341-018-52
3	119.5	9.1	248	4	US-08-470-335-210
4	119.5	9.1	248	4	US-08-470-339-210
5	119.5	9.1	248	4	US-08-467-602-207
6	119.5	9.1	248	4	US-08-467-602-404
7	119.5	9.1	349	4	US-08-470-335-188
8	119.5	9.1	382	4	US-08-467-602-382
9	119.5	9.1	405	4	US-08-467-602-384
10	119.5	9.1	411	4	US-08-470-339-189
11	119.5	9.1	414	4	US-08-470-339-188
12	119.5	9.1	422	1	US-08-036-555B-170
13	119.5	9.1	422	1	US-08-469-569-170
14	119.5	9.1	422	1	US-08-428-926-3
15	119.5	9.1	422	1	US-08-249-322A-170
16	119.5	9.1	422	1	US-08-428-927-3
17	119.5	9.1	422	1	US-08-428-298-3
18	119.5	9.1	422	1	US-08-339-517-3
19	119.5	9.1	422	1	US-08-469-526A-170
20	119.5	9.1	422	2	US-08-734-591A-170
21	119.5	9.1	422	2	US-08-469-660-170
22	119.5	9.1	422	3	US-08-341-018-72
23	119.5	9.1	422	4	US-08-470-335-170
24	119.5	9.1	422	4	US-08-735-021-170
25	119.5	9.1	422	4	US-08-734-664A-170
26	119.5	9.1	422	4	US-08-470-339-170
27	119.5	9.1	422	4	US-08-467-602-170

28	119.5	9.1	422	4	US-08-467-602-324	Sequence 324, App
29	119.5	9.1	422	5	PCT-US94-05083C-166	Sequence 166, App
30	119.5	9.1	422	5	PCT-US94-05083C-185	Sequence 185, App
31	119.5	9.1	422	5	PCT-US95-06846A-170	Sequence 170, App
32	119.5	9.1	425	4	US-08-470-335-226	Sequence 226, App
33	119.5	9.1	425	4	US-08-467-602-320	Sequence 320, App
34	119.5	9.1	445	4	US-08-467-602-328	Sequence 328, App
35	119.5	9.1	456	4	US-08-470-335-246	Sequence 246, App
36	119.5	9.1	456	4	US-08-467-602-303	Sequence 303, App
37	119.5	9.1	456	4	US-08-467-602-366	Sequence 366, App
38	119.5	9.1	459	4	US-08-470-335-239	Sequence 239, App
39	119.5	9.1	459	4	US-08-467-602-299	Sequence 299, App
40	119.5	9.1	459	4	US-08-467-602-362	Sequence 362, App
41	119.5	9.1	479	4	US-08-467-602-307	Sequence 307, App
42	119.5	9.1	479	4	US-08-467-602-370	Sequence 370, App
43	119.5	9.1	490	4	US-08-467-602-345	Sequence 345, App
44	119.5	9.1	493	4	US-08-467-602-341	Sequence 341, App
45	119.5	9.1	513	4	US-08-467-602-349	Sequence 349, App

ALIGNMENTS

RESULT 1
US-09-413-814-78
; Sequence 78, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Beyer, Stefan
; APPLICANT: Bloecker, Helmut
; APPLICANT: Brandt, Petra
; APPLICANT: Cino, Paul M
; APPLICANT: Dougherty, Brian A
; APPLICANT: Goldberg, Steven L
; APPLICANT: Hofle, Gerhard
; APPLICANT: Mueller, Joachim
; APPLICANT: Reichenbach, Hans
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
; FILE REFERENCE: PCT/US 99/23535
; CURRENT APPLICATION NUMBER: US/09/413,814
; CURRENT FILING DATE: 1999-10-07
; EARLIER APPLICATION NUMBER: DE 198 46 493.2
; EARLIER FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 78
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-413-814-78

Query Match 10.4%; Score 137.5; DB 4; Length 882;
Best Local Similarity 27.2%; Pred. No. 0.00031;
Matches 71; Conservative 13; Mismatches 72; Indels 105; Gaps 12;
QY 9 GRAGGVGLLVLLGLFRPPPCALCPVKEPRGL-----SAASPLAETCAPRFRSVP 64
Db 559 GPRGGUG-----ARPGEPRAVRPEPPRAAPPRAGHPAPRRRIHAR 601
QY 65 RGEAGAVQELARALAHLLAEERQARABAEQOQARVLAQLLRVWGAPRNSDPALG 124
Db 602 RA-----PRVRRLVGRRLRRR-----BALRRL-----RAGPAF- 632
QY 125 LDDDPDAPAAQLARALLRLDPAALAAQLVPAPVPAALRPPPVYDD---GPAGPDAE 181
Db 633 -----PAAGAPCAVRRRLRRSPAGVAVRRGPG-----GRAPVLDDALGRAAPGR 677
QY 182 EAGDETPDVDPPELLRYLLGRILAGSAD--SEGVAAPRRLRAADHD-----VGSLEPPEG 234
Db 678 -----AGRPRAGGAKDVARGAAEPRRRGRRAHPPDQGVLRAGGDLPLHG 720

QY 235 VLGALLVRKLETPAPQVPAR 255
Db 721 -----PARRVQAR 728

RESULT 2

US-08-341-018-52
; Sequence 52, Application US/08341018A
; Patent No. 6087323
; GENERAL INFORMATION:
; APPLICANT: Gwynne, David I.
; APPLICANT: Mahanthappa, Nagesh K.
; APPLICANT: Marchionni, Mark A.
; APPLICANT: Birmingham-McDonogh, Olivia
; APPLICANT: Goldin, Stanley M.
; APPLICANT: McBurney, Robert N.
; TITLE OF INVENTION: USE OF NEUREGULINS AS MODULATORS OF
; CELLULAR COMMUNICATION
; FILE REFERENCE: 04585/041001
; CURRENT APPLICATION NUMBER: US/08/341.018A
; CURRENT FILING DATE: 1994-11-17
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 52
; LENGTH: 248
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-341-018-52

Query Match 9.1%; Score 119.5; DB 3; Length 248;
Best Local Similarity 28.0%; Pred. No. 0.0026;
Matches 56; Conservative 15; Mismatches 66; Indels 63; Gaps 9;

QY 2 AGSPLLWGPAGGVG-----LLVLLLLGLFRPPPCALCARPVKEPRGLSA--ASP 48
Db 10 SGRP---GPRQRPASAARSPPPLPLLLLLGLTAAALPAGAAAGNEAAPAGASVCYSSP 66
QY 49 PLAETGAPRRFRSVPGEAAGVQELARALAHLEAEQERARAEQAEEDQQARVLAQ 108
Db 67 P-----SVGSVQELAQRAAVVIEGKVHPQRR--QQGALDRKAAAG 106
QY 109 LLRVGAPRNSDPALGLDDDPAPAAQLARALLRRLDPAALAAQLVPAPVPAALRPRP 168
Db 107 EAGAWG-----DREPPAA-----GPRALGP---PAEPLLAANGTV 140
QY 169 PVYDDGAPGDAEEAGDETP 188
Db 141 PSWPTAPV-PSAGEGEEAP 159

RESULT 3

US-08-470-335-210
; Sequence 210, Application US/08470335F
; Patent No. 6147190
; GENERAL INFORMATION:
; APPLICANT: GOODEARL, ANDREW
; APPLICANT: STROOBANT, PAUL
; APPLICANT: MINGHETTI, LUISA
; APPLICANT: WATERFIELD, MICHAEL
; APPLICANT: MARCHIONNI, MARK
; APPLICANT: CHEN, MARIO S.
; TITLE OF INVENTION: GLIAL MITOGENIC FACTORS, THEIR
; PREPARATION AND USE
; FILE REFERENCE: 04585/00200B
; CURRENT APPLICATION NUMBER: US/08/470.335F
; CURRENT FILING DATE: 1993-03-24
; EARLIER FILING DATE: 1993-03-24
; NUMBER OF SEQ ID NOS: 252
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 210

; LENGTH: 248
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-470-335-210

Query Match 9.1%; Score 119.5; DB 4; Length 248;
Best Local Similarity 28.0%; Pred. No. 0.0026;
Matches 56; Conservative 15; Mismatches 66; Indels 63; Gaps 9;
QY 2 AGSPLLWGPAGGVG-----LLVLLLLGLFRPPPCALCARPVKEPRGLSA--ASP 48
Db 10 SGRP---GPRQRPASAARSPPPLPLLLLLGLTAAALPAGAAAGNEAAPAGASVCYSSP 66
QY 49 PLAETGAPRRFRSVPGEAAGVQELARALAHLEAEQERARAEQAEEDQQARVLAQ 108
Db 67 P-----SVGSVQELAQRAAVVIEGKVHPQRR--QQGALDRKAAAG 106
QY 109 LLRVGAPRNSDPALGLDDDPAPAAQLARALLRRLDPAALAAQLVPAPVPAALRPRP 168
Db 107 EAGAWG-----DREPPAA-----GPRALGP---PAEPLLAANGTV 140
QY 169 PVYDDGAPGDAEEAGDETP 188
Db 141 PSWPTAPV-PSAGEGEEAP 159

RESULT 4

US-08-470-339-210
; Sequence 210, Application US/08470339C
; Patent No. 6232286
; GENERAL INFORMATION:
; APPLICANT: GOODEARL, ANDREW
; APPLICANT: STROOBANT, PAUL
; APPLICANT: MINGHETTI, LUISA
; APPLICANT: WATERFIELD, MICHAEL
; APPLICANT: MARCHIONNI, MARK
; APPLICANT: CHEN, MARIO S.
; APPLICANT: HILES, IAN
; TITLE OF INVENTION: GLIAL MITOGENIC FACTORS, THEIR
; PREPARATION AND USE
; FILE REFERENCE: 04585/00200B
; CURRENT APPLICATION NUMBER: US/08/470.339C
; CURRENT FILING DATE: 1995-06-06
; EARLIER FILING DATE: 1993-03-24
; EARLIER FILING DATE: 1993-03-24
; EARLIER FILING DATE: 1992-09-03
; EARLIER FILING DATE: 1992-06-30
; EARLIER FILING DATE: 1992-06-30
; EARLIER FILING DATE: 1992-04-03
; EARLIER FILING DATE: 1999-04-10
; NUMBER OF SEQ ID NOS: 226
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 210
; LENGTH: 248
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-470-339-210

Query Match 9.1%; Score 119.5; DB 4; Length 248;
Best Local Similarity 28.0%; Pred. No. 0.0026;
Matches 56; Conservative 15; Mismatches 66; Indels 63; Gaps 9;

QY 2 AGSPLLWGPAGGVG-----LLVLLLLGLFRPPPCALCARPVKEPRGLSA--ASP 48
Db 10 SGRP---GPRQRPASAARSPPPLPLLLLLGLTAAALPAGAAAGNEAAPAGASVCYSSP 66
QY 49 PLAETGAPRRFRSVPGEAAGVQELARALAHLEAEQERARAEQAEEDQQARVLAQ 108
Db 67 P-----SVGSVQELAQRAAVVIEGKVHPQRR--QQGALDRKAAAG 106

QY 109 LLRWGAPRNSDPALGLDDPDPAQAQALRALRLRDLPAALAAQLVPAPVPAALRPRP 168
Db 107 EAGAWGG-----DREPPAA-----GPRALGP---PAEPLLAANGTV 140
QY 169 PVYDDGPGAPDAEAGDETP 188
Db 141 PSWPTAPV-PSAGEPGEAP 159

RESULT 5

US-08-467-602-207
; Sequence 207, Application US/08467602C
; Patent No. 6444642
; GENERAL INFORMATION:
; APPLICANT: Sklar, Robert
; APPLICANT: Marchionni, Mark
; APPLICANT: Gwynne, David I.
; TITLE OF INVENTION: METHODS FOR TREATING MUSCLE DISEASES AND
; TITLE OF INVENTION: DISORDERS
; FILE REFERENCE: 04585/028003
; CURRENT APPLICATION NUMBER: US/08/467,602C
; EARLIER FILING DATE: 1995-06-06
; EARLIER APPLICATION NUMBER: 08/209,204
; EARLIER FILING DATE: 1994-03-08
; EARLIER APPLICATION NUMBER: 08/059,022
; EARLIER FILING DATE: 1993-05-06
; NUMBER OF SEQ ID NOS: 420
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 207
; LENGTH: 248
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-467-602-207

Query Match 9.1%; Score 119.5; DB 4; Length 248;
Best Local Similarity 28.0%; Pred. No. 0.0026;
Matches 56; Conservative 15; Mismatches 66; Indels 63; Gaps 9;

QY 2 AGSPLLWGPRAGVG-----LLVLLLLGLFRPPPPALCARPVKEPRGLSA--ASP 48
Db 10 SGRP---GPRAPRGSAAARSPPPLPLLLLLGLTAAALPAGAAAGNEAPAGASVCYSSP 66
QY 49 PLAETGAPRRFRSVPGEAGAVQELARALAHLEAERERARAEAEEDQOARVLAQ 108
Db 67 P-----SVGSVQELAQRAANVIEGKVHPQRR--QQGALDRKAAAAG 106
QY 109 LLRWGAPRNSDPALGLDDPDPAQAQALRALRLRDLPAALAAQLVPAPVPAALRPRP 168
Db 107 EAGAWGG-----DREPPAA-----GPRALGP---PAEPLLAANGTV 140
QY 169 PVYDDGPGAPDAEAGDETP 188
Db 141 PSWPTAPV-PSAGEPGEAP 159

RESULT 6

US-08-467-602-404
; Sequence 404, Application US/08467602C
; Patent No. 6444642
; GENERAL INFORMATION:
; APPLICANT: Sklar, Robert
; APPLICANT: Marchionni, Mark
; APPLICANT: Gwynne, David I.
; TITLE OF INVENTION: METHODS FOR TREATING MUSCLE DISEASES AND
; TITLE OF INVENTION: DISORDERS
; FILE REFERENCE: 04585/028003
; CURRENT APPLICATION NUMBER: US/08/467,602C
; EARLIER FILING DATE: 1995-06-06
; EARLIER APPLICATION NUMBER: 08/209,204
; EARLIER FILING DATE: 1994-03-08
; EARLIER APPLICATION NUMBER: 08/059,022
; NUMBER OF SEQ ID NOS: 420

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 404
; LENGTH: 248
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-467-602-404

Query Match 9.1%; Score 119.5; DB 4; Length 248;
Best Local Similarity 28.0%; Pred. No. 0.0026;
Matches 56; Conservative 15; Mismatches 66; Indels 63; Gaps 9;

QY 2 AGSPLLWGPRAGVG-----LLVLLLLGLFRPPPPALCARPVKEPRGLSA--ASP 48
Db 10 SGRP---GPRAPRGSAAARSPPPLPLLLLLGLTAAALPAGAAAGNEAPAGASVCYSSP 66
QY 49 PLAETGAPRRFRSVPGEAGAVQELARALAHLEAERERARAEAEEDQOARVLAQ 108
Db 67 P-----SVGSVQELAQRAANVIEGKVHPQRR--QQGALDRKAAAAG 106
QY 109 LLRWGAPRNSDPALGLDDPDPAQAQALRALRLRDLPAALAAQLVPAPVPAALRPRP 168
Db 107 EAGAWGG-----DREPPAA-----GPRALGP---PAEPLLAANGTV 140
QY 169 PVYDDGPGAPDAEAGDETP 188
Db 141 PSWPTAPV-PSAGEPGEAP 159

RESULT 7

US-08-470-335-188
; Sequence 188, Application US/08470335F
; Patent No. 6147190
; GENERAL INFORMATION:
; APPLICANT: GOODEARL, ANDREW
; APPLICANT: STROEBANT, PAUL
; APPLICANT: MINCHETTI, LUISA
; APPLICANT: WATERFIELD, MICHAEL
; APPLICANT: MARCHIONNI, MARK
; APPLICANT: CHEN, MARIO S.
; APPLICANT: HILES, IAN
; TITLE OF INVENTION: GLIAL MITOGENIC FACTORS, THEIR
; TITLE OF INVENTION: PREPARATION AND USE
; FILE REFERENCE: 04585/00200B
; CURRENT APPLICATION NUMBER: US/08/470,335F
; EARLIER FILING DATE: 1995-06-06
; EARLIER APPLICATION NUMBER: 08/036,555
; EARLIER FILING DATE: 1993-03-24
; NUMBER OF SEQ ID NOS: 252
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 188
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-470-335-188

Query Match 9.1%; Score 119.5; DB 4; Length 349;
Best Local Similarity 28.0%; Pred. No. 0.004;
Matches 56; Conservative 15; Mismatches 66; Indels 63; Gaps 9;

QY 2 AGSPLLWGPRAGVG-----LLVLLLLGLFRPPPPALCARPVKEPRGLSA--ASP 48
Db 10 SGRP---GPRAPRGSAAARSPPPLPLLLLLGLTAAALPAGAAAGNEAPAGASVCYSSP 66
QY 49 PLAETGAPRRFRSVPGEAGAVQELARALAHLEAERERARAEAEEDQOARVLAQ 108
Db 67 P-----SVGSVQELAQRAANVIEGKVHPQRR--QQGALDRKAAAAG 106
QY 109 LLRWGAPRNSDPALGLDDPDPAQAQALRALRLRDLPAALAAQLVPAPVPAALRPRP 168
Db 107 EAGAWGG-----DREPPAA-----GPRALGP---PAEPLLAANGTV 140
QY 169 PVYDDGPGAPDAEAGDETP 188

Db 141 PSWPTAPV-PSAGEPGEEAP 159

RESULT 8

US-08-467-602-382
; Sequence 382, Application US/08467602C
; Patent No. 6444642

```

; GENERAL INFORMATION:
;
; APPLICANT: Sklar, Robert
;
; APPLICANT: Marchionni, Mark
;
; APPLICANT: Gwynne, David I.
;
; TITLE OF INVENTION: METHODS FOR TREATING MUSCLE DISEASES AND
;
; TITLE OF INVENTION: DISORDERS
;
; FILE REFERENCE: 04585/028003
;
; CURRENT APPLICATION NUMBER: US/08/467,602C
;
; CURRENT FILING DATE: 1995-06-06
;
; EARLIER APPLICATION NUMBER: 08/209,204
;
; EARLIER FILING DATE: 1994-03-08
;
; EARLIER APPLICATION NUMBER: 08/059,022
;
; EARLIER FILING DATE: 1993-05-06
;
; NUMBER OF SEQ ID NOS: 420
;
; SOFTWARE: FastSeq for Windows Version 4.0
;
; SEQ ID NO 382
;
; LENGTH: 382
;

```

TYPE: PRT

ORGANISM: *Homo sapiens*

; FEATURE:

NAME/KEY: VARIANT

NAME/REF: VARIANI
LOCATION: (34) (34)

LOCATION: (34)...(34)
OTHER INFORMATION: Xaa is any amino acid

US-08-467-602-382

Query Match 9.18: Score 119.5; DB 4: Length 382;

Best Local Similarity 28.08; Pred. No. 0.0045;

Matches	56;	Conservative	15;	Mismatches	66;	Indels	63
---------	-----	--------------	-----	------------	-----	--------	----

QY 2 AGSPLLWGPRAGVG-----LLVLLLLGLFRPPPALCARPVKEPRGLSA--ASP 48

[illegible]

Db 44 SGRP--GPRORPGSAARSSPPLPLPLLLLGTAALAPGAAAGNEAAPAGASVCYSSP 100

[illegible]

QV 49 PLAETGAPRRFRSVPRGEAAGAVOELARALAHLLLEAEROERARAEAOEAEDOOARVLAQ 108

[illegible]

Db 101 P-----SVGSVOELAORAAVWTEGKVHPORR--OOGATDRKAAAAAG 140

DD 101 F SVGSVQEERQNRNVIEGRVHFQRR - QQGALDNRNNAHTAG 140

0v 109 I.I.RVWGAPRNSDPAIGLDDDDPDAPAAOIA.RA.I.I.RARL.DPAALAAOI.VPAPVPAAAI.PPPP 168

QY IUV LLRVWGAEKNSDFATGLGDDDPDAPPAQ LARALLKRALDLPAPALAAQLVFPAPVFAAALRPFR 168

141 FACAWCC-----DBEPDPA-----CDBATCO-----DREEDTANCTMY 174

DU 141 EAGAWG-----DREPPA-----GPRALGP-----PAEPLLAANGIV I/4

0" 160 BIVVDCBACDBACEEAE 100

QY 169 P VYDDGFPAGDAEEAGDETP 188

שנה	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100																																																																																																																																																																																									
175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	4

DB I/5 PSWPTAPV-PSAGEPGEAP 193

SECRET

US-08-467-602-384
; Sequence 384, Application US/08467602C
; Patent No. 6444642

```

1  / SEQUENCE NO. 0444042
2  /
3  / GENERAL INFORMATION:
4  /
5  / APPLICANT: Sklar, Robert
6  /
7  / APPLICANT: Marchionni, Mark
8  /
9  / APPLICANT: Gwynne, David I.
10 /
11 / TITLE OF INVENTION: METHODS FOR TREATING MUSCLE DISEASES AND
12 /
13 / TITLE OF INVENTION: DISORDERS
14 /
15 / FILE REFERENCE: 04585/028003
16 /
17 / CURRENT APPLICATION NUMBER: US/08/467,602C
18 /
19 / CURRENT FILING DATE: 1995-06-06
20 /
21 / EARLIER APPLICATION NUMBER: 08/209,204
22 /
23 / EARLIER FILING DATE: 1994-03-08
24 /
25 / EARLIER APPLICATION NUMBER: 08/059,022
26 /
27 / EARLIER FILING DATE: 1993-05-06
28 /
29 / NUMBER OF SEQ ID NOS: 420
30 /
31 / SOFTWARE: FastSeq for Windows Version 4.0
32 /
33 / SEQ ID NO 384
34 /

```

; SEQ ID NO 384

• 2

QY 49 PLAETGAPRRFRSVPGEAGAVOELARALAHLLLEAERQERARAEAEEDQOARVLAQ 108
Db 67 P-----SVGSVOELAQRAAAVVEGKVPQRR--QOQALDRKAAAAG 106
QY 109 LLRWGAPRNSDPALGLDDPDAPAAQALARALLRARLDPAALAAQLVPAPVPAALRPRP 168
Db 107 EAGAWGG-----DREPPAA-----GPRALGP---PAEPLLAANGTV 140

QY 169 PVYDDGPGAPDAEAGDETP 188

Db 141 PSWPTAPV-PSAGEPGEAP 159

RESULT 11

US-08-470-339-188

; Sequence 188, Application US/08470339C

; Patent No. 6232286

; GENERAL INFORMATION:

; APPLICANT: GOODEARL, ANDREW

; APPLICANT: STROOBANT, PAUL

; APPLICANT: MINGHETTI, LUISA

; APPLICANT: WATERFIELD, MICHAEL

; APPLICANT: MARCHIONNI, MARK

; APPLICANT: CHEN, MARIO S.

; APPLICANT: HILES, IAN

; TITLE OF INVENTION: GLIAL MITOGENIC FACTORS, THEIR

; FILE REFERENCE: 04585/002008

; CURRENT APPLICATION NUMBER: US/08/470,339C

; CURRENT FILING DATE: 1995-06-06

; EARLIER APPLICATION NUMBER: 08/036,555

; EARLIER FILING DATE: 1993-03-24

; EARLIER APPLICATION NUMBER: 07/940,389

; EARLIER FILING DATE: 1992-09-03

; EARLIER APPLICATION NUMBER: 07/907,138

; EARLIER FILING DATE: 1992-06-30

; EARLIER APPLICATION NUMBER: 07/863,703

; EARLIER FILING DATE: 1992-04-03

; EARLIER APPLICATION NUMBER: 91 07566.3 GB

; EARLIER FILING DATE: 1999-04-10

; NUMBER OF SEQ ID NOS: 226

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 188

; LENGTH: 414

; TYPE: PRT

; ORGANISM: Homo sapiens

; US-08-470-339-188

Query Match 9.1%; Score 119.5; DB 4; Length 414;

Best Local Similarity 28.0%; Pred. No. 0.0049;

Matches 56; Conservative 15; Mismatches 66; Indels 63; Gaps 9;

QY 2 AGSPLLMGPRAGGVG-----LLVLILLGLFRPPPPALCARPVKEPRGLSA--ASP 48

Db 10 SGRP--GPRAGRGSARSPPPLPLPLLLLTALAPGAAAGNEAPAGASVCYSSP 66

QY 49 PLAETGAPRRFRSVPGEAGAVOELARALAHLLLEAERQERARAEAEEDQOARVLAQ 108

Db 67 P-----SVGSVOELAQRAAAVVEGKVPQRR--QOQALDRKAAAAG 106

QY 109 LLRWGAPRNSDPALGLDDPDAPAAQALARALLRARLDPAALAAQLVPAPVPAALRPRP 168

Db 107 EAGAWGG-----DREPPAA-----GPRALGP---PAEPLLAANGTV 140

QY 169 PVYDDGPGAPDAEAGDETP 188

Db 141 PSWPTAPV-PSAGEPGEAP 159

RESULT 12

US-08-036-555B-170

; Sequence 170, Application US/08036555B

; Patent No. 5530109

GENERAL INFORMATION:

; APPLICANT: Goodearl, Andrew; Stroobant, Paul;

; APPLICANT: Minghetti, Luisa; Waterfield, Michael; Marchioni, Mark;

; APPLICANT: Chen, Maio Su; Hiles, Ian

; TITLE OF INVENTION: Glial Mitogenic Factors, Their

; TITLE OF INVENTION: Preparation and Use

; NUMBER OF SEQUENCES: 184

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Felfe & Lynch

; STREET: 805 Third Avenue

; CITY: New York City

; STATE: New York

; COUNTRY: USA

; ZIP: 10022

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage

; COMPUTER: IBM

; OPERATING SYSTEM: PC-DOS

; SOFTWARE: Wordperfect

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/036,555B

; FILING DATE: 24-MAR-1993

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/965,173

; FILING DATE: 23-OCT-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/940,389

; FILING DATE: 03-SEP-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/907,138

; FILING DATE: 30-JUN-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/863,703

; FILING DATE: 03-APRIL-1992

; APPLICATION DATA:

; APPLICATION NUMBER: U.K. 91 07566.3

; FILING DATE: 10-APRIL-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: Tsai, Christine H.

; REGISTRATION NUMBER: 34,266

; REFERENCE/DOCKET NUMBER: LUD 5250.4

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 688-9200

; TELEFAX: (212) 838-3884

; INFORMATION FOR SEQ ID NO: 170:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 422

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: linear

; US-08-036-555B-170

Query Match 9.1%; Score 119.5; DB 1; Length 422;

Best Local Similarity 28.0%; Pred. No. 0.0051;

Matches 56; Conservative 15; Mismatches 66; Indels 63; Gaps 9;

QY 2 AGSPLLMGPRAGGVG-----LLVLILLGLFRPPPPALCARPVKEPRGLSA--ASP 48

Db 10 SGRP--GPRAGRGSARSPPPLPLPLLLLTALAPGAAAGNEAPAGASVCYSSP 66

QY 49 PLAETGAPRRFRSVPGEAGAVOELARALAHLLLEAERQERARAEAEEDQOARVLAQ 108

Db 67 P-----SVGSVOELAQRAAAVVEGKVPQRR--QOQALDRKAAAAG 106

QY 109 LLRWGAPRNSDPALGLDDPDAPAAQALARALLRARLDPAALAAQLVPAPVPAALRPRP 168

Db 107 EAGAWGG-----DREPPAA-----GPRALGP---PAEPLLAANGTV 140

QY 169 PVYDDGPGAPDAEAGDETP 188

Db 141 PSWPTAPV-PSAGEPGEAP 159

RESULT 15

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2003, 14:18:20 ; Search time 36 Seconds
(without alignments)
441.538 Million cell updates/sec

Title: US-09-803-589-6

Perfect score: 1319

Sequence: 1 MAGSPLLMGPRAGGVGLLV.....RVKRLTPAPQVPRLLPP 260

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 248812 seqs, 61136040 residues

Total number of hits satisfying chosen parameters: 248812

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1319	100.0	260	10	US-09-803-589-6
2	627	47.5	175	10	US-09-803-589-14
3	308	23.4	149	9	US-09-984-271-150
4	308	23.4	166	9	US-09-984-271-238
5	124.5	9.4	418	9	US-09-946-807-3
6	124.5	9.4	418	10	US-09-795-668-3
7	124.5	9.4	418	10	US-09-795-686-3
8	118.5	9.0	550	9	US-09-976-740-47
9	118.5	9.0	550	12	US-10-023-529-47
10	118.5	9.0	550	12	US-10-023-523-47
11	118.5	9.0	1298	9	US-09-825-288A-2
12	114.5	8.7	492	10	US-09-764-898-206
13	112	8.5	2630	9	US-10-077-130-2
14	112	8.5	7968	9	US-10-077-130-5
15	110.5	8.4	538	12	US-09-976-740-43
16	110.5	8.4	538	12	US-10-023-529-43
17	110.5	8.4	538	12	US-10-023-523-43
18	109	8.3	681	10	US-09-815-242-11830
19	109	8.3	1665	10	US-09-858-664A-2

Sequence 3, Appli
Sequence 216, App
Sequence 216, App
Sequence 5, Appli
Sequence 8, Appli
Sequence 2, Appli
Sequence 1037, Ap
Sequence 283, App
Sequence 100, App
Sequence 333, App
Sequence 331, App
Sequence 3, Appli
Sequence 2, Appli
Sequence 4, Appli
Sequence 6, Appli
Sequence 8, Appli
Sequence 425, App
Sequence 425, App
Sequence 425, App
Sequence 425, App
Sequence 394, App
Sequence 394, App
Sequence 394, App
Sequence 394, App
Sequence 394, App
Sequence 394, App

20 108 8.2 846 10 US-09-858-664A-3
21 107.5 8.2 527 9 US-09-854-133-216
22 107.5 8.2 527 10 US-09-738-973-216
23 106 8.0 7257 9 US-10-014-717-5
24 105.5 8.0 1213 10 US-09-969-515-8
25 105.5 8.0 1224 10 US-09-969-515-2
26 103 8.0 404 10 US-09-764-864-1037
27 103 8.0 439 9 US-10-097-065-283
28 105 8.0 578 10 US-09-888-615-100
29 104 7.9 649 9 US-09-759-1308-333
30 104 7.9 671 9 US-09-759-1308-331
31 104 7.9 827 9 US-10-171-384-3
32 103.5 7.8 811 10 US-09-768-436-2
33 103.5 7.8 965 10 US-09-737-149-4
34 103.5 7.8 971 10 US-09-737-149-6
35 103.5 7.8 1043 10 US-09-737-149-8
36 103.5 7.8 1184 9 US-09-978-295A-425
37 103.5 7.8 1184 9 US-09-978-697-425
38 103.5 7.8 1184 9 US-09-978-192A-425
39 103.5 7.8 1184 9 US-09-999-832A-425
40 103.5 7.8 1184 9 US-09-978-189-425
41 103.5 7.8 1184 9 US-10-028-072-394
42 103.5 7.8 1184 9 US-10-121-049-394
43 103.5 7.8 1184 9 US-10-123-904-394
44 103.5 7.8 1184 9 US-10-140-470-394
45 103.5 7.8 1184 9 US-10-175-746-394

ALIGNMENTS

RESULT 1

US-09-803-589-6

; Sequence 6, Application US/09803589

; Patent No. US20020112251A1

; GENERAL INFORMATION:

; APPLICANT: McCarthy, Sean A.

; APPLICANT: Holtzman, Douglas A.

; APPLICANT: Goodearl, Andrew D.J.

; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING

; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC AND OTHER

; TITLE OF INVENTION: US6

; FILE REFERENCE: 07334-325001

; CURRENT APPLICATION NUMBER: US/09/803,589

; CURRENT FILING DATE: 2001-03-09

; PRIOR APPLICATION NUMBER: US 09/128,709

; PRIOR FILING DATE: 1998-08-04

; PRIOR APPLICATION NUMBER: US 60/054,645

; PRIOR FILING DATE: 1997-08-04

; PRIOR APPLICATION NUMBER: US 09/130,491

; PRIOR FILING DATE: 1998-08-06

; PRIOR APPLICATION NUMBER: US 60/054,966

; PRIOR FILING DATE: 1997-08-06

; PRIOR APPLICATION NUMBER: US 60/058,108

; PRIOR FILING DATE: 1997-09-05

; PRIOR APPLICATION NUMBER: US 09/388,280

; PRIOR FILING DATE: 1999-09-01

; PRIOR APPLICATION NUMBER: US 09/388,279

; PRIOR FILING DATE: 1999-09-01

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 6

; LENGTH: 260

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-803-589-6

Query Match 100.0%; Score 1319; DB 10; Length 260;
Best Local Similarity 100.0%; Pred. No. 9.9e-87;
Matches 260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGSPLLMGPRAGGVGLLVLLLLLGLFRPPPPALCARPVKEPGLSAASPPLAETGAPRRFR 60

|||||

```
Db 1 MAGSPLWGPRAGGVGLLVLLGLFRPPPPALCARPVKEPRGLSAASPPPLAETGAPRRFR 60
QY 61 RSVPRGEAAGAVQELARALAHLEAEQERARAEAEQEDQARVLAQLLRVWGAPRNSD 120
Db 61 RSVPRGEAAGAVQELARALAHLEAEQERARAEAEQEDQARVLAQLLRVWGAPRNSD 120
QY 121 PALGLDDDDPAPAAQARALLRRALDPAALAAQLVPAAPVPAALRRPPVPYDDGAPGPD 180
Db 121 PALGLDDDDPAPAAQARALLRRALDPAALAAQLVPAAPVPAALRRPPVPYDDGAPGPD 180
QY 181 EAGDFTPDVDPPELLRYLLGRILAGSADSEGVAAAPRRRLRAADHDVGSLEPPEGVLGALL 240
Db 181 EAGDFTPDVDPPELLRYLLGRILAGSADSEGVAAAPRRRLRAADHDVGSLEPPEGVLGALL 240
QY 241 RVKLETPAPQVPARRLLPP 260
Db 241 RVKLETPAPQVPARRLLPP 260

RESULT 2
US-09-803-589-14
; Sequence 14, Application US/09803589
; Patent No. US20020112251A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Goodearl, Andrew D.J.
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC AND OTHER
; TITLE OF INVENTION: USES
; FILE REFERENCE: 07334-325001
; CURRENT APPLICATION NUMBER: US/09/803,589
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: US 09/128,709
; PRIOR FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: US 60/054,645
; PRIOR FILING DATE: 1997-08-04
; PRIOR APPLICATION NUMBER: US 09/130,491
; PRIOR FILING DATE: 1998-08-06
; PRIOR APPLICATION NUMBER: US 60/054,966
; PRIOR FILING DATE: 1997-08-06
; PRIOR APPLICATION NUMBER: US 60/058,108
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: US 09/388,280
; PRIOR FILING DATE: 1999-09-01
; PRIOR APPLICATION NUMBER: US 09/388,279
; PRIOR FILING DATE: 1999-09-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 175
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-803-589-14

Query Match 47.5%; Score 627; DB 10; Length 175;
Best Local Similarity 78.0%; Pred. No. 8.6e-38;
Matches 131; Conservative 6; Mismatches 27; Indels 4; Gaps 1;

QY 1 MAGSPLWGPRAGGVGLLVLLGLFRPPPPALCARPVKEPRGLSAASPPPLAETGAPRRFR 60
Db 1 MAGSPLWGPRAGGVGLLVLLGLFRPPPPALCARPVKEPRGLSAASAPLVETSTPLRLR 60
QY 61 RSVPRGEAAGAVQELARALAHLEAEQERARAEAEQEDQARVLAQLLRVWGAPRNSD 120
Db 61 RAVPRGEAAGAVQELARALAHLEAEQERARAEAEQEDQARVLAQLLRWAGSPRASD 120
QY 121 PALGLDDDDPAPAAQARALLRRALDPAALAAQLVPAAPVPAALRRPPVPYDDGAPGPD 164
Db 121 PPLAPDDDDPAPAAQARALLRRALDPAALAAQLVPAAPVPAALRRPPVPYDDGAPGPD 168

RESULT 3
```

```
US-09-984-271-150
; Sequence 150, Application US/09984271
; Publication No. US20030040088A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 71 Human Secreted Proteins
; FILE REFERENCE: PZ030P1
; CURRENT APPLICATION NUMBER: US/09/984,271
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 09/482,273
; PRIOR FILING DATE: 2000-01-13
; PRIOR APPLICATION NUMBER: PCT/US99/15849
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: 60/092,921
; PRIOR FILING DATE: 1998-07-15
; PRIOR APPLICATION NUMBER: 60/092,922
; PRIOR FILING DATE: 1998-07-15
; PRIOR APPLICATION NUMBER: 60/092,956
; PRIOR FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 267
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 150
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (149)
; OTHER INFORMATION: xaa equals stop translation
US-09-984-271-150

Query Match 23.4%; Score 308; DB 9; Length 149;
Best Local Similarity 50.3%; Pred. No. 3.2e-15;
Matches 88; Conservative 3; Mismatches 34; Indels 50; Gaps 8;

QY 1 MAGSPLWGPRAGGVGLLVLLGLFRPPPPALCARPVKEPRGLSAASPPPLAETG-----54
Db 1 MAGSPLWGPRAGGVGLLVLLGLFRPPPPALCARPVKEPRGLSAASPPPLARLALAAG 60
QY 55 --APRRFRSVPR-GEAAGAVQELARALAHLEAEQERARAEAEQEDQARVLAQLLR 111
Db 61 GQCPVRRRGRCRPGAGAGA-----SAGAEQERARAEQNR-----LR 98
QY 112 V-----WGAPRNSDPALGLDDDDPAPAAQLAR-----ALLRARLDPAALAAQLVPA 157
Db 99 ISRRASW-----RSCCAGS-----APPATLIRLWAMTTTPTLRQSSALCSAPA 143

RESULT 4
US-09-984-271-238
; Sequence 238, Application US/09984271
; Publication No. US20030040088A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 71 Human Secreted Proteins
; FILE REFERENCE: PZ030P1
; CURRENT APPLICATION NUMBER: US/09/984,271
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 09/482,273
; PRIOR FILING DATE: 2000-01-13
; PRIOR APPLICATION NUMBER: PCT/US99/15849
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: 60/092,921
; PRIOR FILING DATE: 1998-07-15
; PRIOR APPLICATION NUMBER: 60/092,922
; PRIOR FILING DATE: 1998-07-15
; PRIOR APPLICATION NUMBER: 60/092,956
; PRIOR FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 267
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 238
; LENGTH: 166
; TYPE: PRT
```

```
; ORGANISM: Homo sapiens
US-09-984-271-238

Query Match      23.4%; Score 308; DB 9; Length 166;
Best Local Similarity 50.3%; Pred. No. 3.6e-15;
Matches 88; Conservative 3; Mismatches 34; Indels 50; Gaps 8;

QY 1 MAGSPILWGPAGGVLVLLVLLGLFRPPPPALCARPVKEPRGLSAAAPPLAETG----- 54
Db 19 MAGSPILWGPAGGVLVLLVLLGLFRPPPPALCARPVKEPRGLSAAAPPLARLALLAASG 78
QY 55 --APRRFRFRSVPR-GEAAGAVQELARALAHLLAEQERARAEAEQAEQDQARVLAQLLR 111
Db 79 GQCEVRRRCRGCGAGAGA-----SAGAEQERARAEAQ-----LR 116
QY 112 V-----WGAPRNSDPALGLDDDDPDAPAAQLAR-----ALLRARLDPAALAAQLVPA 157
Db 117 ISRRASW-----RSCCAGS-----APPATLRLWAWTTTPTLRQSSSLALCSAPA 161

RESULT 5
US-09-946-807-3
; Sequence 3, Application US/09946807
; Patent No. US20020165144A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinhorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2004-001
; CURRENT APPLICATION NUMBER: US/09/946,807
; CURRENT FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: US/09/795,668
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/515,716
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-946-807-3

Query Match      9.4%; Score 124.5; DB 9; Length 418;
Best Local Similarity 28.8%; Pred. No. 0.11;
Matches 57; Conservative 14; Mismatches 64; Indels 63; Gaps 9;

QY 2 AGSPLLWGPAGGVLVLLVLLGLFRPPPPALCARPVKEPRGLSAAAPPL 50
Db 10 SGRP---GPRARPGSAARSPPPLPLLLGLLGTAAAPGAA--AGNEAAPAGASVCSP- 63
QY 51 AETGAPRRFRSVPRGEAAGAVQELARALAHLLAEQERARAEAEQAEQDQARVLAQLL 110
Db 64 -----SVGSVQELAQRAAVVIEGKVHPQRR--QQGALDRKAAAGAA 104
QY 111 RVWGAPRNSDPALGLDDDDPDAPAAQLARLRLARLDPAALAAQLVPAPVPAALRPPPV 170
Db 105 GAWGG-----DREPPAA-----GPRALGP---PAEEPLLAANGTVPS 138
QY 171 YDDGPGADPAEEAGDETP 188
Db 139 WPTAPV-PSAGEPGEAP 155

RESULT 6
US-09-946-807-3
; Sequence 3, Application US/09946807
; Patent No. US20020165144A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinhorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2004-001
; CURRENT APPLICATION NUMBER: US/09/946,807
; CURRENT FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: US/09/795,668
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/515,716
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-946-807-3

Query Match      9.4%; Score 124.5; DB 9; Length 418;
Best Local Similarity 28.8%; Pred. No. 0.11;
Matches 57; Conservative 14; Mismatches 64; Indels 63; Gaps 9;

QY 2 AGSPLLWGPAGGVLVLLVLLGLFRPPPPALCARPVKEPRGLSAAAPPL 50
Db 10 SGRP---GPRARPGSAARSPPPLPLLLGLLGTAAAPGAA--AGNEAAPAGASVCSP- 63
QY 51 AETGAPRRFRSVPRGEAAGAVQELARALAHLLAEQERARAEAEQAEQDQARVLAQLL 110
Db 64 -----SVGSVQELAQRAAVVIEGKVHPQRR--QQGALDRKAAAGAA 104
QY 111 RVWGAPRNSDPALGLDDDDPDAPAAQLARLRLARLDPAALAAQLVPAPVPAALRPPPV 170
Db 105 GAWGG-----DREPPAA-----GPRALGP---PAEEPLLAANGTVPS 138
QY 171 YDDGPGADPAEEAGDETP 188
Db 139 WPTAPV-PSAGEPGEAP 155

RESULT 6
US-09-946-807-3
; Sequence 3, Application US/09946807
; Patent No. US20020165144A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinhorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
```

```
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2004-001
; CURRENT APPLICATION NUMBER: US/09/795,668
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/515,716
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-795-668-3

Query Match      9.4%; Score 124.5; DB 10; Length 418;
Best Local Similarity 28.8%; Pred. No. 0.11;
Matches 57; Conservative 14; Mismatches 64; Indels 63; Gaps 9;

QY 2 AGSPLLWGPAGGVLVLLVLLGLFRPPPPALCARPVKEPRGLSAAAPPL 50
Db 10 SGRP---GPRARPGSAARSPPPLPLLLGLLGTAAAPGAA--AGNEAAPAGASVCSP- 63
QY 51 AETGAPRRFRSVPRGEAAGAVQELARALAHLLAEQERARAEAEQAEQDQARVLAQLL 110
Db 64 -----SVGSVQELAQRAAVVIEGKVHPQRR--QQGALDRKAAAGAA 104
QY 111 RVWGAPRNSDPALGLDDDDPDAPAAQLARLRLARLDPAALAAQLVPAPVPAALRPPPV 170
Db 105 GAWGG-----DREPPAA-----GPRALGP---PAEEPLLAANGTVPS 138
QY 171 YDDGPGADPAEEAGDETP 188
Db 139 WPTAPV-PSAGEPGEAP 155

RESULT 7
US-09-795-686-3
; Sequence 3, Application US/09795686
; Patent No. US20020094954A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinhorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2005-001
; CURRENT APPLICATION NUMBER: US/09/795,686
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/515,715
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-795-686-3

Query Match      9.4%; Score 124.5; DB 10; Length 418;
Best Local Similarity 28.8%; Pred. No. 0.11;
Matches 57; Conservative 14; Mismatches 64; Indels 63; Gaps 9;

QY 2 AGSPLLWGPAGGVLVLLVLLGLFRPPPPALCARPVKEPRGLSAAAPPL 50
Db 10 SGRP---GPRARPGSAARSPPPLPLLLGLLGTAAAPGAA--AGNEAAPAGASVCSP- 63
QY 51 AETGAPRRFRSVPRGEAAGAVQELARALAHLLAEQERARAEAEQAEQDQARVLAQLL 110
Db 64 -----SVGSVQELAQRAAVVIEGKVHPQRR--QQGALDRKAAAGAA 104
QY 111 RVWGAPRNSDPALGLDDDDPDAPAAQLARLRLARLDPAALAAQLVPAPVPAALRPPPV 170
Db 105 GAWGG-----DREPPAA-----GPRALGP---PAEEPLLAANGTVPS 138
```



```
; SEQ ID NO 47
; LENGTH: 550
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
US-10-023-523-47

Query Match          9.0%; Score 118.5; DB 12; Length 550;
Best Local Similarity 28.4%; Pred. No. 0.39;
Matches 76; Conservative 14; Mismatches 103; Indels 75; Gaps 12;

QY 10 PRAGGVGGLVLLGLFRPPALCARPVKEPRG--LSAASF-----PLAETGAP 56
Db 103 PRGAT-----PPA-----PPRPRGGPAAAAAPPTTAPPPPPAPVAAAAP 145
QY 57 RFRFRSVPRGEAGAVOELARALAHLEAEQERARAEQAEEDQARVLAQLLRVWGAP 116
Db 146 AR-----APRAAAAAA-----ATAPSPGPAQPGPRA-----QRAAPLAAPPPAPAAP 189
QY 117 RNSDPALGLDDDDPAAPAAQIAR-ALLRARLDPAALAAQLVPAPVAAAALRPRPPVYDDGP 175
Db 190 PAAAPAGPRAPPPAAVAARESPLPPPPQPPAPQQQQQQPPPPPPPPQPP-----P 244
QY 176 AGPDAEAGDETDPDPELLRYLLGRILGASADSEG-----VAAPRRLRRAADHDV 226
Db 245 EGGAARAGGPAPVSVLRVYVILGGSGAGGRLTRGRVQGLLEEAARGLERTR----- 300
QY 227 GSELPPEVGLALLRVKRLTETAPQVPA 254
Db 301 -----LGA-LALPRGDRPGRAPPA 318

RESULT 11
US-09-825-288A-2
; Sequence 2, Application US/09825288A
; Publication No. US20020192822A1
; GENERAL INFORMATION:
; APPLICANT: ROIZMAN, BERNARD
; TITLE OF INVENTION: HERPES SIMPLEX VIRUS ICP4 IS AN INHIBITOR OF APOPTOSIS
; FILE REFERENCE: ARCD:317USC1
; CURRENT APPLICATION NUMBER: US/09/825,288A
; PRIOR FILING DATE: 2001-04-02
; PRIOR FILING DATE: 1999-03-01
; PRIOR FILING DATE: 1996-07-26
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1298
; TYPE: PRT
; ORGANISM: HERPES VIRUS, TYPE 1
US-09-825-288A-2

Query Match          9.0%; Score 118.5; DB 9; Length 1298;
Best Local Similarity 24.9%; Pred. No. 1;
Matches 88; Conservative 24; Mismatches 121; Indels 121; Gaps 15;

QY 2 AGSP-LIWGPRAGVGGLVLLGLFRPPALCARPVKEPR----- 41
Db 355 SGAPAAVWAEPLGDAAQQAQYALITRLTYTPDAEAMGWLONRPPVPGDVALDQACFRISGAA 414
QY 42 -----GLSAASPLAETGAPRFRRSVPGEAGAVQ-----ELARAL 79
Db 415 RNSSSFTTGSVARVPHLGYMAAGRFGWGLAHAAAVAMRRYDRAQKGFLLTSLRRAY 474
QY 80 AHLEAER-----QERARAEQAEQO-----ARVLAQLLRVWGAP 116
Db 475 APLLARENAALTGAAGSPGAGADDEGVAAVAAAAPGERAVPAGYGAAGIILALGRLSAAP 534
QY 117 RNSDPALGLDDDDPAP-----AAQLAR-----ALLRARLDPAALAAQLVPAPV 159
Db 535 --ASPAGG--DDPDAARHADDDAGRRAGQAGRAVVECLAACRGILEALAEFGDGLAAV 590

; SEQ ID NO 160
; LENGTH: 202
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
US-10-077-130-2

Query Match          8.7%; Score 114.5; DB 10; Length 492;
Best Local Similarity 25.1%; Pred. No. 0.66;
Matches 78; Conservative 19; Mismatches 87; Indels 127; Gaps 16;

QY 160 PA-AALRPRPPVYDDGPGAGPDAEAGDETDPDPELLRYLLGRI----- 202
Db 591 PGLAGARPASPPRPEGPAGP-----ASPPPHADAPRLRAWLRELFVRDALVLMRLRGDL 646
QY 203 -LAGSADSEGVAAAPRRLRRAADHDVGSSELPEGVIG-ALLRVKRLTETAPQVPA 254
Db 647 RVAGGSEA-AVAAVRAVSLVA-----GALGPALPRDPRLPSSAAAAA 688

RESULT 12
US-09-764-898-206
; Sequence 206, Application US/09764898
; Patent No. US20020090673A1
; GENERAL INFORMATION:
; APPLICANT: ROSEN ET AL.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ201
; CURRENT APPLICATION NUMBER: US/09/764,898
; CURRENT FILING DATE: 2001-01-17
; PRIOR APPLICATION DATA REMOVED - CONSULT PALM OR FILE WRAPPER
; NUMBER OF SEQ ID NOS: 311
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 206
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-898-206

Query Match          8.7%; Score 114.5; DB 10; Length 492;
Best Local Similarity 25.1%; Pred. No. 0.66;
Matches 78; Conservative 19; Mismatches 87; Indels 127; Gaps 16;

QY 28 PPPALCARPVKEPRGLSAAFP-----PLAETGAPRFR-----RS 62
Db 194 PPEDAPAP-----PTPGPAASPEQLSFRERKYELEVVRVQAE-GPKRVSLVGCADDLRK 249
QY 63 VPGEAGAGVQELARALAHLEAEQERARAE-----QEAEDQARVLAQLLRVWGAPR 117
Db 250 MQEEEARKLQKQKRAQMLREAAEAGAEARLADGETTGLTEEQEQEQP-----PWASPS 301
QY 118 -----NSDPALG-----LDDDDPAALAAQLARALLRA----- 143
Db 302 PTSRQSPASPPPLGGGAPVRTAKARRHOERLRVQSPPEPPAPERALSPEALRALEAEKRA 361
QY 144 -----RLDPAALAAQLV-----PAPVPAALRPRP-PVYDDG 174
Db 362 LWRAARMKSLEQDALRAQWVLSRQSGRGTRGPLERLAEAPSPAPT-----PSTPTVEDLG 417
QY 175 P-----AGPD-AEEAGDETDPDVP-----ELLRYLLGRILGASADSEGVAA-APR 216
Db 418 PQTSTSPGRSLSPDFAELRSLEFSPSPGQEDGEVALVLLGRSPGAVGPDVALCSSLR 477
QY 217 RLRRADHDVG 227
Db 478 RPVRPGRRGGLG 488

RESULT 13
US-10-077-130-2
; Sequence 2, Application US/10077130
; Patent No. US20020168742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family
; FILE REFERENCE: MPI2001-047P1RCPI(M)
; CURRENT APPLICATION NUMBER: US/10/077,130
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269201
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 2630
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-077-130-2

Query Match 8.5%; Score 112; DB 9; Length 2630;
Best Local Similarity 27.7%; Pred. No. 6.4;
Matches 81; Conservative 23; Mismatches 98; Indels 90; Gaps 15;

QY 28 PPPALCARP-VKEPRGLSASPPL---AETGAPRRFRSVPGEAAGAV----- 72
DB 1804 PGPSLDAEGTQEAEDLSDTPTLQRPQEQATMRKFSLG-GRGGYAGVAGYGTFAFGDA 1862
QY 73 -----QELARALAHLEAERQARAEQAQAEQQAQVLAQLLRVWG----- 114
DB 1863 GGMGLQGQPMWARIWAVSQ-SEEEQEAEARAESEEQEAEAEPLQVVSARPPVEVGR 1921
QY 115 APRNSDPA-----LG-----LDDDPDAPAAQALARALLRRLDPAAL----- 150
DB 1922 APTRSSPEPTWEDIGQVSLQIRDLSDGDAEAAADI-SLIDISEVDPAYLNLSLDYDIKYL 1980
QY 151 -----AAQLVPAPVPA-----ALRPPVPYDDGAGPADAE-EAGDETDPVDPE 193
DB 1981 PFEWIFRKVPKSAQPEP-PSMAEEELAEPEPTWMPWPGELGPHAGLEITESESDVDAL 2039
QY 194 LLRYLLGRILAGSADSEGVAAAPRRRLRAADHDVGSSELPE--GVLGALLRVK 243
DB 2040 LAEAAVGRKRKWSPPS-----RSLFHPGRRHLPLDEPAELGLRERVK 2081

RESULT 14
US-10-077-130-5
; Sequence 5, Application US/10077130
; Patent No. US20020168742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family
; FILE REFERENCE: MPI2001-047P1RCPl(M)
; CURRENT APPLICATION NUMBER: US/10/077,130
; PRIOR FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269201
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 7968
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-077-130-5

Query Match 8.5%; Score 112; DB 9; Length 7968;
Best Local Similarity 27.7%; Pred. No. 22;
Matches 81; Conservative 23; Mismatches 98; Indels 90; Gaps 15;

QY 28 PPPALCARP-VKEPRGLSASPPL---AETGAPRRFRSVPGEAAGAV----- 72
DB 7142 PGPSLDAEGTQEAEDLSDTPTLQRPQEQATMRKFSLG-GRGGYAGVAGYGTFAFGDA 7200
QY 73 -----QELARALAHLEAERQARAEQAQAEQQAQVLAQLLRVWG----- 114
DB 7201 GGMGLQGQPMWARIWAVSQ-SEEEQEAEARAESEEQEAEAEPLQVVSARPPVEVGR 7259
QY 115 APRNSDPA-----LG-----LDDDPDAPAAQALARALLRRLDPAAL----- 150
DB 7260 APTRSSPEPTWEDIGQVSLQIRDLSDGDAEAAADI-SLIDISEVDPAYLNLSLDYDIKYL 7318
QY 151 -----AAQLVPAPVPA-----ALRPPVPYDDGAGPADAE-EAGDETDPVDPE 193
DB 7319 PFEWIFRKVPKSAQPEP-PSMAEEELAEPEPTWMPWPGELGPHAGLEITESESDVDAL 7377

QY 194 LLRYLLGRILAGSADSEGVAAAPRRRLRAADHDVGSSELPE--GVLGALLRVK 243
DB 7378 LAEAAVGRKRKWSPPS-----RSLFHPGRRHLPLDEPAELGLRERVK 7419

RESULT 15
US-09-976-740-43
; Sequence 43, Application US/09976740
; Publication No. US20020194633A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/976,740
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 43
; LENGTH: 538
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-740-43

Query Match 8.4%; Score 110.5; DB 9; Length 538;
Best Local Similarity 26.7%; Pred. No. 1.4;
Matches 77; Conservative 17; Mismatches 99; Indels 95; Gaps 16;

QY 1 MAGSLLMGPRAGGVGLLVLLLLGLFRPPPPALCARPVKEPRGLSASP-----LAETGA 55
DB 1 MAGPPAL-----PPETATAAATTAASSSSAASPHYQEWILDTID 40
QY 56 PRFRFRSVPGENAGAVQELARALAHLEAERQARAEQAQAEQQAQVLAQLLRV--- 112
DB 41 SLRSRKARP-----DLERICR-WVRRRHGPEPEPTRAELEKLIQRA-----VLRSYK 88
QY 113 -----WGAPRNSDPALGLDDDDPDAP--AAQLARALLRRLDPAALAAQLVPAPVPA--A 163
DB 89 GSISYRNAARVQPPRRGA--TPPAPPAPRAGAPAAAAAAPPPTAPPAPVAAAAAPA 146
QY 164 LRPR-----PPVYDDGPA--GPDAAEAGD-ETPDVDPELLRYLLGRILAGSADSEGV 212
DB 147 RAPRAAAAAATAPP--SPGPAQGPRAQAAPLAAPPAP-----AAPPAPV 190
QY 213 AAPRLRRAADHDVGSSELPEGVLGALLRVKRLTETPAQVQVARRLLPP 260
DB 191 APPAGPRAPPVAAAREPP-----LPPPPQPPA-----PP 221

Search completed: April 4, 2003, 14:26:42
Job time : 42 secs


```
XX PF 26-DEC-2000; 2000WO-US34263.
XX PR 21-JAN-2000; 2000US-0488725.
XX PR 25-APR-2000; 2000US-0523217.
XX PR 09-JUL-2000; 2000US-0598042.
XX PR 19-JUL-2000; 2000US-0620312.
XX PR 03-AUG-2000; 2000US-0653450.
XX PR 14-SEP-2000; 2000US-0662191.
XX PR 19-OCT-2000; 2000US-0693036.
XX PR 29-NOV-2000; 2000US-0727344.
XX PA (HYSE-) HYSEQ INC.
XX OS
XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX XX
XX DR WPI: 2001-442253/47.
XX DR N-PSDB; AA158161.
XX XX
XX XX Novel nucleic acids and polypeptides, useful for treating disorders
XX PT such as central nervous system injuries -
XX PT
XX PS Example 4; SEQ ID NO 2150; 10078pp; English.
XX XX
XX CC The invention relates to human nucleic acids (AA157798-AA161369) and
XX CC the encoded polypeptides (AAAM38642-AAAM42213) with nootropic,
XX CC immunosuppressant and cytostatic activity. The polynucleotides are useful
XX CC in gene therapy. A composition containing a polypeptide or polynucleotide
XX CC of the invention may be used to treat diseases of the peripheral nervous
XX CC system, such as peripheral nervous injuries, peripheral neuropathy and
XX CC localised neuropathies and central nervous system diseases, such as
XX CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX CC utilisation of the activities such as: Immune system suppression,
XX CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX CC assays for receptor activity, arthritis and inflammation, leukaemias and
XX CC C.N.S disorders.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification.
XX XX
XX XX Sequence 260 AA;
XX XX
XX XX Query Match 100.0%; Score 260; DB 22; Length 260;
XX XX Best Local Similarity 100.0%; Pred. No. 7.8e-213;
XX XX Matches 260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX QY 1 MAGSPLLWGPRAGGVGLVLLGLLGLRPPPPALCARVPKPRGLSAASPPPLAETGAPRRFR 60
XX DB 1 MAGSPLLWGPRAGGVGLVLLGLLGLRPPPPALCARVPKPRGLSAASPPPLAETGAPRRFR 60
XX QY 61 RSVPRGENAGAVQELARALAHLEAEQERARAEAEQEQARVLAQLLRVWGAPRNSD 120
XX DB 61 RSVPRGENAGAVQELARALAHLEAEQERARAEAEQEQARVLAQLLRVWGAPRNSD 120
XX QY 121 PALGLDDDDPAPAAQALARALLRLRLDPAALAAQLVPAPVPAALRPPVYDDGPAGPDA 180
XX DB 121 PALGLDDDDPAPAAQALARALLRLRLDPAALAAQLVPAPVPAALRPPVYDDGPAGPDA 180
XX QY 181 EAGDETPDVPDPELLRYLLGRILAGSADSEGVAAAPRRRLRAADHDVGSLEPPGVLGALL 240
XX DB 181 EAGDETPDVPDPELLRYLLGRILAGSADSEGVAAAPRRRLRAADHDVGSLEPPGVLGALL 240
XX QY 241 RVKRLETPAPQVPAARLLPP 260
XX DB 241 RVKRLETPAPQVPAARLLPP 260
XX XX
XX XX RESULT 4
XX XX AAB88596
XX ID AAB88596 standard; Protein; 260 AA.
XX XX
XX AC AAB88596;
XX XX
XX DT 04-JUN-2001 (first entry)
XX DE Human hydrophobic domain containing protein clone HP10756 #100.
XX XX
XX KW Human; hydrophobic domain; immunosuppressant; anti-HIV; neuroprotective;
XX KW antanaemic; vulnery; antiulcer; osteopathic; anti-inflammatory;
XX KW cytostatic; gene therapy; autoimmune disorder; multiple sclerosis;
XX KW HIV infection; anaemia; burn; ulcer; osteoporosis; tumour; wound healing;
XX KW inflammatory bowel disease; nutritional supplement; appetite; vaccine;
XX KW behavioural characteristic; immune response.
XX OS
XX OS Homo sapiens.
XX PN WO200112660-A2.
XX XX
XX PD 22-FEB-2001.
XX XX
XX PF 10-AUG-2000; 2000WO-JP05356.
XX PR 17-AUG-1999; 99JP-0230344.
XX PR 07-SEP-1999; 99JP-0252551.
XX PR 01-OCT-1999; 99JP-0281132.
XX PR 22-OCT-1999; 99JP-0301624.
XX PR 04-NOV-1999; 99JP-0313877.
XX XX
XX PA (SAGA ) SAGAMI CHEM RES CENT.
XX PA (PROT-) PROTEGENE INC.
XX PI Kato S, Kimura T;
XX XX
XX DR WPI; 2001-160059/16.
XX DR N-PSDB; AAF94486.
XX XX
XX PT Human proteins with hydrophobic domains and the DNAs which encode them
XX PT are useful for treating autoimmune disorders, burns and tumors and for
XX PT screening novel pharmaceuticals -
XX XX
XX PS Claim 1; Page 422-423; 518pp; English.
XX XX
XX CC AAF94417 to AAF94516 encode the human proteins given in AAB88557 to
XX CC AAB88606 (I) which have a hydrophobic domain. (I) have immunosuppressant,
XX CC anti-HIV, neuroprotective, antianaemic, vulnery, antiulcer,
XX CC osteopathic, anti-inflammatory and cytostatic activities, and can be
XX CC used in gene therapy. (I) can be used as pharmaceuticals and as antigens
XX CC to prepare antibodies. DNA and cDNA (II) encoding (I) can be used as
XX CC probes for genetic diagnosis and gene sources for gene therapy or for
XX CC producing (I) in large quantities. Cells containing (II) are used for
XX CC the detection of ligands or receptors corresponding to membrane or
XX CC secretory proteins and to screen small molecule novel pharmaceuticals.
XX CC Antibodies directed to (I) can be used for the detection, quantification
XX CC and purification of (I). Activities of (I) may include cytokine and cell
XX CC proliferation/differentiation function, immune stimulating or suppressing
XX CC activity, haematopoiesis regulating activity, tissue growth activity,
XX CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX CC and thrombolytic activity, receptor/ligand activity and anti-inflammatory
XX CC activity. (I) and (II) can be used to treat autoimmune disorders e.g.
XX CC multiple sclerosis, HIV infections, anaemia, burns, ulcers, osteoporosis,
XX CC inflammatory bowel disease and tumors. (I) and (II) can also be used for
XX CC wound healing, as nutritional sources or supplements e.g. as amino acid,
XX CC carbon or nitrogen source, to effect metabolism, catabolism, anabolism,
XX CC processing and utilisation of dietary fat, protein, carbohydrate,
XX CC vitamins and minerals, to effect behavioural characteristics, to affect
XX CC appetite, and can act as antigens in vaccines to raise an immune response
XX CC to the protein or another material cross-reactive with the protein.
XX XX
XX SQ Sequence 260 AA;
XX XX
XX XX Query Match 100.0%; Score 260; DB 22; Length 260;
XX XX Best Local Similarity 100.0%; Pred. No. 7.8e-213;
XX XX Matches 260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```


XX (HUMA-) HUMAN GENOME SCI INC.
 XX Ruben SM, Komatsoulis GA, Duan DR, Rosen CA, Moore PA, Shi Y;
 PI Lafleur DW, Olsen HS, Brewer LA, Florence KA, Young PE, Soppet DR;
 PI Endress GA, Mucenski M, Ebner R;
 XX WPI; 2001-425865/45.
 DR N-PSDB; AAD11677.
 XX Isolated nucleic acid molecule encoding a human secreted protein is
 PT used in preventing, treating or ameliorating a medical condition -
 XX Claim 11; Page 773-774; 864pp; English.
 XX AAD11630-AAD11721 represent cDNAs corresponding to 71 human secreted
 CC protein genes, and AAE06041-AAE06132 represent the proteins they encode.
 CC AAE06133-AAE06205 represent human secreted protein fragments.
 CC The secreted proteins and their genes are useful for preventing, treating
 CC or ameliorating medical conditions, e.g., by protein or gene therapy.
 CC Pathological conditions can be diagnosed by determining the amount of the
 CC new protein in a sample or by determining the presence of mutations in
 CC the new genes. Specific uses are described for each of the 71 genes,
 CC based on the tissues in which they are most highly expressed, and include
 CC developing products for the diagnosis or treatment of proliferative
 CC disorders, cancer, tumours, foetal and developmental abnormalities,
 CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune
 CC diseases (e.g., rheumatoid arthritis), inflammation, allergies,
 CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),
 CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,
 CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,
 CC angenogenic disorders, kidney disorders, gastrointestinal disorders,
 CC pregnancy-related disorders, endocrine disorders, and infections. The
 CC proteins can also be used to aid wound healing and epithelial cell
 CC proliferation, to prevent skin aging due to sunburn, to maintain organs
 CC before transplantation, for supporting cell culture of primary tissues,
 CC to regenerate tissues, to identify their cognate ligands or binding
 CC partners, and in chemotaxis, and can be used as a food additive or
 CC preservative to modify storage properties. Antibodies specific for a
 CC protein of the invention can be used in alleviating symptoms associated
 CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,
 CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA).
 CC The present sequence represents a human secreted protein of
 CC the invention.
 XX Sequence 148 AA;
 SQ Query Match. 19.6%; Score 51; DB 22; Length 148;
 Best Local Similarity 100.0%; Pred. No. 2e-35;
 Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MAGSPLLWGPAGGVLVLLGLFRPPPPALCARPVKEPRGLSAASPPLA 51
 DB 1 MAGSPLLWGPAGGVLVLLGLFRPPPPALCARPVKEPRGLSAASPPLA 51
 RESULT 7
 ABG33910
 ID ABG33910 standard; Protein; 148 AA.
 XX AC ABG33910;
 XX DT 15-JUL-2002 (first entry)
 XX DE Human secreted protein encode by gene 48.
 XX Human: secreted protein; gene therapy; immunosuppressive;
 KW antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
 KW vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
 KW virucide; fungicide; opthalmaloidal; autoimmune disease; neoplasm;
 KW rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;
 KW cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;
 KW angiogenesis; nervous system disorder; Alzheimer's disease; infection;

KW ocular disorder; corneal infection; wound healing; skin aging;
 KW epithelial cell proliferation; food additive.
 OS Homo sapiens.
 PN WO200226931-A2.
 XX 04-APR-2002.
 PD 24-SEP-2001; 2001WO-US29871.
 XX 25-SEP-2000; 2000US-234925P.
 PR 12-JAN-2001; 2001WO-US00911.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Ruben SM, Komatsoulis G, Duan DR, Rosen CA, Moore PA, Shi Y;
 PI Lafleur DW, Olsen H, Brewer LA, Florence KA, Young PE, Soppet DR;
 PI Endress GA, Mucenski M, Ebner R;
 XX WPI; 2002-362489/39.
 DR N-PSDB; ABK69773.
 XX Novel 71 isolated secreted polypeptides and polynucleotides encoding
 PT the polypeptides, useful for treating Huntington's disease, sepsis,
 PT meningitis, thrombocytopaenia, haemolytic anaemia, rheumatoid arthritis,
 PT asthma
 XX Claim 11; Page 1252-1253; 1478pp; English.
 XX The invention relates to an isolated nucleic acid molecule (or its
 CC fragment, homologue complement or allelic variant) encoding a human
 CC secreted protein (and its fragment, domain, epitope, variant, secreted
 CC form and species variant). Also included are a recombinant vector
 CC comprising the nucleic acid, a recombinant host cell comprising the
 CC vector, an antibody against the secreted protein, a recombinant host cell
 CC that expresses the secreted protein and a method of identifying a binding
 CC partner of the secreted protein. The nucleic acid and protein are used to
 CC prevent, diagnose, treat or ameliorate a medical condition in e.g.
 CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep
 CC for example autoimmune diseases e.g. rheumatoid arthritis,
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
 CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi and
 CC ocular disorders e.g. corneal infection. Many other diseases and
 CC disorders are listed in the specification. The polypeptides can also be
 CC used to aid wound healing an epithelial cell proliferation, to prevent
 CC skin aging due to sunburn, to maintain organs before transplantation, for
 CC supporting cell culture of primary tissues, to regenerate tissues and in
 CC chemotaxis. The polypeptides can also be used as a food additive or
 CC preservative to increase or decrease storage capabilities. The present
 CC sequence represents a novel human secreted protein of the invention.
 XX Sequence 148 AA;
 SQ Query Match. 19.6%; Score 51; DB 23; Length 148;
 Best Local Similarity 100.0%; Pred. No. 2e-35;
 Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MAGSPLLWGPAGGVLVLLGLFRPPPPALCARPVKEPRGLSAASPPLA 51
 DB 1 MAGSPLLWGPAGGVLVLLGLFRPPPPALCARPVKEPRGLSAASPPLA 51
 RESULT 8
 AAY87111
 ID AAY87111 standard; Protein; 149 AA.
 XX AC AAY87111;
 XX DT 09-MAY-2000 (first entry)
 XX

DE Human secreted protein sequence SEQ ID NO:150.

XX Human; secreted protein; diagnosis; cytostatic; cytotoxic; immunosuppressive;
KW antiinflammatory; nontropic; neuroprotective; antiallergic; cancer;
KW tumour; neurodegenerative disorder; developmental abnormality; allergy;
KW foetal deficiency; blood disorder; immune system disorder; arthritis;
KW autoimmune disease; hepatic disease; renal disease; inflammation;
KW Alzheimer's disease; behavioural disorder; schizophrenia; osteoporosis;
KW infection; AIDS; spinal cord injury; transplant rejection; diabetes;
KW asthma; sepsis; acne; psoriasis; cardiovascular disorder;
KW reproductive disorder; gastrointestinal disorder; respiratory disorder;
KW metabolic disorder; food additive; preservative.

XX Homo sapiens.

OS WO200004140-A1.

XX 27-JAN-2000.

XX 14-JUL-1999; 99WO-US15849.

XX 15-JUL-1998; 98US-0092921.

XX 15-JUL-1998; 98US-0092922.

XX 15-JUL-1998; 98US-0092956.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Komatsoulis G, Duan RD, Rosen CA, Moore PA, Shi Y;
PI Lafleur DW, Ebner R, Olsen HS, Brewer LA, Florence KA, Young PE;
PI Mucenski M, Endress GA, Soppet DR;

XX WPI: 2000-161128/14.

DR N-PSDB; AAZ98064.

PT New isolated human genes, useful for diagnosis and treatment of, e.g.
PT cancers, neurological or blood disorders

XX Claim 11; Page 406-407; 494pp; English.

XX The polynucleotide sequences given in AAZ98017 to AAZ98108 encode the
CC human secreted proteins given in AAZ97064 to AAZ97223. Human secreted
CC protein can have activities based on the tissues and cells the genes are
CC expressed in. Examples of activities include: cytostatic;
CC immunosuppressive; antiinflammatory; nontropic; neuroprotective; and
CC antiallergic. The polynucleotides and their corresponding secreted
CC polypeptides are useful for preventing, treating or ameliorating medical
CC conditions, e.g. by protein or gene therapy. Also pathological conditions
CC can be diagnosed by determining the amount of the new polypeptides in a
CC sample or by determining the presence of mutations in the new
CC polynucleotides. Human secreted protein s and their polynucleotides can
CC be used for developing products for the diagnosis or treatment of cancer,
CC tumours, neurodegenerative disorders, developmental abnormalities and
CC foetal deficiencies, blood disorders, diseases of the immune system,
CC autoimmune diseases, hepatic and renal disease, inflammation,
CC allergies, Alzheimer's disease, behavioural disorders, schizophrenia,
CC osteoporosis, arthritis, infections, AIDS, spinal cord injuries,
CC transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,
CC cardiovascular disorders, reproductive disorders, gastrointestinal
CC disorders, respiratory disorders and metabolic disorders. The
CC proteins or polynucleotides can also be used as food additives or
CC preservatives. The proteins are also useful for identifying their
CC binding partners. AAZ98008 to AAZ98016 and AAZ97063 are sequence used in
CC the exemplification of the present invention.

XX Sequence 149 AA;

SQ Query Match 19.6%; Score 51; DB 21; Length 149;
Best Local Similarity 100.0%; Pred. No. 2e-35;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGSPLWGPAGGVLVLLGLFLRPPALCARPVKEPRGLSAAGPPLA 51
|||||
DB 1 MAGSPLWGPAGGVLVLLGLFLRPPALCARPVKEPRGLSAAGPPLA 51

RESULT 9
ID AAZ97199 standard; Protein; 166 AA.
XX
AC AAZ97199;
XX
DT 09-MAY-2000 (first entry)
XX
DE Human secreted protein sequence SEQ ID NO:238.
XX
KW Human; secreted protein; diagnosis; cytostatic; cytotoxic; immunosuppressive;
KW antiinflammatory; nontropic; neuroprotective; antiallergic; cancer;
KW tumour; neurodegenerative disorder; developmental abnormality; allergy;
KW foetal deficiency; blood disorder; immune system disorder; arthritis;
KW autoimmune disease; hepatic disease; renal disease; inflammation;
KW Alzheimer's disease; behavioural disorder; schizophrenia; osteoporosis;
KW infection; AIDS; spinal cord injury; transplant rejection; diabetes;
KW asthma; sepsis; acne; psoriasis; cardiovascular disorder;
KW reproductive disorder; gastrointestinal disorder; respiratory disorder;
KW metabolic disorder; food additive; preservative.

XX Homo sapiens.

OS WO200004140-A1.

XX 27-JAN-2000.

XX 14-JUL-1999; 99WO-US15849.

XX 15-JUL-1998; 98US-0092921.

XX 15-JUL-1998; 98US-0092922.

XX 15-JUL-1998; 98US-0092956.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Komatsoulis G, Duan RD, Rosen CA, Moore PA, Shi Y;
PI Lafleur DW, Ebner R, Olsen HS, Brewer LA, Florence KA, Young PE;
PI Mucenski M, Endress GA, Soppet DR;

XX WPI: 2000-161128/14.

DR N-PSDB; AAZ98064.

PT New isolated human genes, useful for diagnosis and treatment of, e.g.
PT cancers, neurological or blood disorders

XX Claim 11; Page 470; 494pp; English.

XX The polynucleotide sequences given in AAZ98017 to AAZ98108 encode the
CC human secreted proteins given in AAZ97064 to AAZ97223. Human secreted
CC protein can have activities based on the tissues and cells the genes are
CC expressed in. Examples of activities include: cytostatic;
CC immunosuppressive; antiinflammatory; nontropic; neuroprotective; and
CC antiallergic. The polynucleotides and their corresponding secreted
CC polypeptides are useful for preventing, treating or ameliorating medical
CC conditions, e.g. by protein or gene therapy. Also pathological conditions
CC can be diagnosed by determining the amount of the new polypeptides in a
CC sample or by determining the presence of mutations in the new
CC polynucleotides. Human secreted protein s and their polynucleotides can
CC be used for developing products for the diagnosis or treatment of cancer,
CC tumours, neurodegenerative disorders, developmental abnormalities and
CC foetal deficiencies, blood disorders, diseases of the immune system,
CC autoimmune diseases, hepatic and renal disease, inflammation,
CC allergies, Alzheimer's disease, behavioural disorders, schizophrenia,
CC osteoporosis, arthritis, infections, AIDS, spinal cord injuries,
CC transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,
CC cardiovascular disorders, reproductive disorders, gastrointestinal
CC disorders, respiratory disorders and metabolic disorders. The
CC proteins or polynucleotides can also be used as food additives or
CC preservatives. The proteins are also useful for identifying their
CC binding partners. AAZ98008 to AAZ98016 and AAZ97063 are sequence used in
CC the exemplification of the present invention.

[illegible]

secreted protein (and its fragment, domain, epitope, variant, secreted form and species variant). Also included are a recombinant vector comprising the nucleic acid, a recombinant host cell comprising the vector, an antibody against the secreted protein, a recombinant host cell that expresses the secreted protein and a method of identifying a binding partner of the secreted protein. The nucleic acid and protein are used to prevent, diagnose, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep for example autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection. Many other diseases and disorders are listed in the specification. The polypeptides can also be used to aid wound healing an epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities. The present sequence represents a novel human secreted protein of the invention.

Query Match 19.6%; Score 51; DB 23; Length 166;
Best Local Similarity 100.0%; Pred. No. 2.2e-35;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAGSPLLWGPAGGVLLVLLGLFRPPPCALCARVKEPRGLSASPPLA 51
DB 19 MAGSPLLWGPAGGVLLVLLGLFRPPPCALCARVKEPRGLSASPPLA 69

RESULT 12
AAV71960
ID AAY71960 standard; Protein; 258 AA.
XX AAY71960;
XX 26-MAR-2001 (first entry)
XX Mouse pituitary hormone, pituitrone.
XX Mouse; pituitary; pituitrone; therapy; immune disorder; anaemia;
KW Digeorge syndrome; haematopoietic cell; ataxia telangiectasia;
KW Wiskott-Aldrich disorder; blood coagulation disorder; Addison's disease;
KW autoimmune disorder; multiple sclerosis; systemic lupus erythematosus;
KW SLE; hyperproliferative disorder; gene therapy; neoplasm;
KW infectious disease; immunomodulatory; cytostatic; antimicrobial.
XX Mus sp.
XX Location/Qualifiers
XX Key 1..34
XX Peptide /label= Signal_peptide
XX /note= "An alternative signal sequence referred in
FT page 11"
XX Peptide 1..33
XX /label= Signal_peptide
XX Protein 35..258
XX /note= "Mature mouse pituitrone"
XX /note= "An alternative mature protein referred in
FT page 11"
XX Protein 34..258
XX /note= "Mature mouse pituitrone"
XX WO2000066778-A1.
XX 09-NOV-2000.
XX 27-APR-2000; 2000WO-US11211.
XX 30-APR-1999; 99US-0131966.

(HUMA-) HUMAN GENOME SCI INC.
XX Ruben SM, N1 J;
XX WPI; 2000-687547/67.
XX N-PSDB; AAD02076.
XX Nucleic acids encoding a novel pituitary hormone (pituitrone), useful for diagnosing, preventing and treating e.g. immune disorders, hyperproliferative disorders and blood coagulation disorders -
XX Claim 11; Fig 2; 277pp; English.
XX The present sequence is a mouse pituitary hormone, pituitrone.
XX pituitrone is highly expressed in pituitary gland and is
XX also expressed in brain tissues, spinal cord and kidney. It may be
XX used as antigens in the production of antibodies against pituitrone
XX and in assays to identify modulators. Pituitrone cDNAs are
XX also useful in gene therapy. Pituitrone may be used in the prevention,
XX treatment and diagnosis of diseases associated with inappropriate
XX pituitrone expression. It may be useful in treating disorders related
XX to reproductive and renal system, immune disorders, disorders of
XX haematopoietic cells (e.g., anaemia, Digeorge syndrome, ataxia
XX telangiectasia and Wiskott-Aldrich disorder), blood coagulation
XX disorders, autoimmune disorders (e.g. Addison's disease, multiple
XX sclerosis and systemic lupus erythematosus (SLE)), hyperproliferative
XX disorders (e.g. neoplasms of bone, liver and pancreas) and infectious
XX diseases.
XX Sequence 258 AA;
XX Query Match 12.7%; Score 33; DB 21; Length 258;
XX Best Local Similarity 100.0%; Pred. No. 6.5e-20;
XX Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 126 DDDPDAPAAQLARALLRARLDPAALAAQLVPAP 158
DB 126 DDDPDAPAAQLARALLRARLDPAALAAQLVPAP 158
RESULT 13
AAV71962
ID AAY71962 standard; Protein; 262 AA.
XX AAY71962;
XX 26-MAR-2001 (first entry)
XX Consensus sequence of human, mouse and rat pituitrone.
XX Rat; human; mouse; pituitary; pituitrone; therapy; immune disorder;
KW Digeorge syndrome; haematopoietic cell; ataxia telangiectasia;
KW Wiskott-Aldrich disorder; blood coagulation disorder; Addison's disease;
KW autoimmune disorder; multiple sclerosis; systemic lupus erythematosus;
KW SLE; hyperproliferative disorder; gene therapy; neoplasm; anaemia;
KW infectious disease; immunomodulatory; cytostatic; antimicrobial.
XX Rattus sp.
XX Mus sp.
XX Homo sapiens.
XX Location/Qualifiers
XX Key Misc-difference 8..12
XX /label= Unknown
XX Misc-difference 17
XX /label= Unknown
XX Misc-difference 26
XX /label= Unknown
XX Misc-difference 28
XX /label= Unknown
XX Misc-difference 31
XX /label= Unknown

FT Misc-difference 33 /label= Unknown
FT Misc-difference 42 /label= Unknown
FT Misc-difference 46 /label= Unknown
FT Misc-difference 48 /label= Unknown
FT Misc-difference 51 /label= Unknown
FT Misc-difference 54 /label= Unknown
FT Misc-difference 55 /label= Unknown
FT Misc-difference 57 /label= Unknown
FT Misc-difference 59 /label= Unknown
FT Misc-difference 62 /label= Unknown
FT Misc-difference 79 /label= Unknown
FT Misc-difference 81 /label= Unknown
FT Misc-difference 82 /label= Unknown
FT Misc-difference 108 /label= Unknown
FT Misc-difference 111 /label= Unknown
FT Misc-difference 112 /label= Unknown
FT Misc-difference 115 /label= Unknown
FT Misc-difference 118 /label= Unknown
FT Misc-difference 122 /label= Unknown
FT Misc-difference 124 /label= Unknown
FT Misc-difference 125 /label= Unknown
FT Misc-difference 159 /label= Unknown
FT /label= Unknown
FT /note= "Xaa equals any amino acid residue or no amino acid residue"
FT Misc-difference 160 /label= Unknown
FT /note= "Xaa equals any amino acid residue or no amino acid residue"
FT Misc-difference 164 /label= Unknown
FT Misc-difference 176 /label= Unknown
FT Misc-difference 180 /label= Unknown
FT Misc-difference 182 /label= Unknown
FT Misc-difference 207 /label= Unknown
FT /note= "Xaa equals any amino acid residue or no amino acid residue"
FT Misc-difference 208 /label= Unknown
FT /note= "Xaa equals any amino acid residue or no amino acid residue"
FT Misc-difference 216 /label= Unknown
FT /note= "Xaa equals any amino acid residue or no amino acid residue"
FT Misc-difference 217 /label= Unknown
FT /note= "Xaa equals any amino acid residue or no amino acid residue or no

FT amino acid residue"
XX WO200066778-A1.
XX 09-NOV-2000.
XX 27-APR-2000; 2000WO-US11211.
XX 30-APR-1999; 99US-0131966.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Ruben SM, Ni J;
XX WPI; 2000-687547/67.
XX Nucleic acids encoding a novel pituitary hormone (pituitrone), useful for diagnosing, preventing and treating e.g. immune disorders, hyperproliferative disorders and blood coagulation disorders -
XX Claim 11; Fig 2; 277pp; English.
XX The present sequence is a consensus sequence of human, mouse and rat pituitrone. Pituitrone is highly expressed in pituitary gland and is also expressed in brain tissues, spinal cord and kidney. Pituitrone may be used as antigens in the production of antibodies against pituitrone and in assays to identify modulators. Pituitrone cDNAs are also useful in gene therapy. Pituitrone may be used in the prevention, treatment and diagnosis of diseases associated with inappropriate pituitrone expression. It may be useful in treating disorders related to reproductive and renal system, immune disorders, disorders of haematopoietic cells (e.g., anaemia, Digeorge syndrome, ataxia telangiectasia and Wiskott-Aldrich disorder), blood coagulation disorders, autoimmune disorders (e.g. Addison's disease, multiple sclerosis and systemic lupus erythematosus (SLE)), hyperproliferative disorders (e.g. neoplasms of bone, liver and pancreas) and infectious diseases.
XX SQ Sequence 262 AA;
Query Match 12.7%; Score 33; DB 21; Length 262;
Best Local Similarity 100.0%; Pred. No. 6.6e-20;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 126 DDDPDAPAAQLARALLRARLDPAALAAQLVPAP 158
|||||
DB 126 DDDPDAPAAQLARALLRARLDPAALAAQLVPAP 158
RESULT 14
AA71961
ID AA71961 standard; Protein; 69 AA.
XX AA71961;
XX 26-MAR-2001 (first entry)
XX Rat pituitary hormone, pituitrone.
XX Rat; pituitary; pituitrone; therapy; immune disorder; anaemia; Digeorge syndrome; haematopoietic cell; ataxia telangiectasia; Wiskott-Aldrich disorder; blood coagulation disorder; Addison's disease; autoimmune disorder; multiple sclerosis; systemic lupus erythematosus; SLE; hyperproliferative disorder; gene therapy; neoplasm; infectious disease; immunomodulatory; cytostatic; antimicrobial.
XX Rattus sp.
XX WO200066778-A1.
XX 09-NOV-2000.
XX 27-APR-2000; 2000WO-US11211.

```
XX 30-APR-1999; 99US-0131966.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Ruben SM, N1 J;
XX WPI; 2000-687547/67.
XX N-PSDB; AAD02077.
XX Nucleic acids encoding a novel pituitary hormone (pituitrone), useful
PT for diagnosing, preventing and treating e.g. immune disorders,
PT hyperproliferative disorders and blood coagulation disorders -
XX Claim 11; Fig 2; 277pp; English.
XX The present sequence is a rat pituitary hormone, pituitrone.
CC Pituitrone is highly expressed in pituitary gland and is
CC also expressed in brain tissues, spinal cord and kidney. It may be
CC used as antigens in the production of antibodies against pituitrone
CC and in assays to identify modulators. Pituitrone cDNAs are
CC also useful in gene therapy. Pituitrone may be used in the prevention,
CC treatment and diagnosis of diseases associated with inappropriate
CC pituitrone expression. It may be useful in treating disorders related
CC to reproductive and renal system, immune disorders, disorders of
CC haematopoietic cells (e.g., anaemia, Digorge syndrome, ataxia
CC telangiectasia and Wiskott-Aldrich disorder), blood coagulation
CC disorders, autoimmune disorders (e.g. Addison's disease, multiple
CC sclerosis and systemic lupus erythematosus (SLE)), hyperproliferative
CC disorders (e.g. neoplasms of bone, liver and pancreas) and infectious
CC diseases.
XX SQ Sequence 69 AA;
Query Match 4.6%; Score 12; DB 21; Length 69;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 235 VLGALLRVKRL 246
Db 44 VLGALLRVKRL 55
|||||
```

Search completed: April 4, 2003, 14:28:29
Job time : 74 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 12, 2003, 15:48:41 ; Search time 275 Seconds
(without alignments)
8017.114 Million cell updates/sec

Title: US-09-803-589-5

Perfect score: 979

Sequence: 1 gaattcgacagagccagc.....gaggatccctaccctggc 979

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

- 1: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA2001A.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA2001B.DAT.*
- 24: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	977	99.8	979	20	AAV68668
2	961.4	98.2	1030	22	AAI58161
3	961.4	98.2	1041	22	AAI59947
4	961.4	98.2	1041	22	AAI59948
5	961.4	98.2	1043	21	AAI59948
6	959.8	98.0	998	22	AAI59949
7	912.8	93.2	1050	21	AAZ98064
8	912.8	93.2	1050	22	AAZ98064
9	912.8	93.2	1050	24	ABK69773

10	776.8	79.3	780	22	AAF94486	Human hydrophobic
11	729.4	74.5	919	22	AAI58162	Human polynucleoti
12	565.4	57.8	991	21	AAI58162	cDNA encoding mous
13	474.2	48.4	4503	24	ABK69876	Human secreted pro
14	301.6	30.8	652	24	ABQ31154	Oligonucleotide fo
15	301.6	30.8	652	24	ABQ31155	Oligonucleotide fo
16	291.8	29.8	652	24	ABQ31156	Oligonucleotide fo
17	291.8	29.8	652	24	ABQ31157	Oligonucleotide fo
18	276.6	28.3	301	24	ABN95203	Gene #1701 used to
19	276.6	28.3	301	24	ABL68387	Kidney cancer rela
20	153.2	15.6	396	21	AAI58162	cDNA encoding rat
21	99.6	10.2	114955	20	AAI58162	Human adenosine A1
22	98.8	10.1	739	24	ABQ31160	Oligonucleotide fo
23	98.8	10.1	739	24	ABQ31161	Oligonucleotide fo
24	95.2	9.7	114955	20	AAI58162	Human adenosine A1
25	90.6	9.3	739	24	ABQ31158	Oligonucleotide fo
26	90.6	9.3	739	24	ABQ31159	Oligonucleotide fo
27	77.8	7.9	4403765	22	AAI99683	Mycobacterium tube
28	75.2	7.7	1614	22	AAH26499	Human low density
29	75.2	7.7	1614	22	AAH26499	Human low density
30	73.4	7.5	2561	22	AAH26500	Rabbit low density
31	71.6	7.3	4403765	22	AAI99683	Mycobacterium tube
32	70.4	7.2	3957	22	AAI99686	HSV-2 immediate ea
33	70.4	7.2	154746	24	AAI99686	Human herpesvirus
34	70.4	7.2	154746	24	AAI99686	Human herpesvirus
35	69.6	7.1	888	17	AAI59269	Streptomyces prist
36	69.6	7.1	2888	17	AAI59268	Streptomyces prist
37	69.4	7.1	2307	24	ABK32842	DNA encoding human
38	69.4	7.1	3957	22	AAI99686	HSV-2 immediate ea
39	69.4	7.1	4411529	22	AAI99682	Mycobacterium tube
40	68.8	7.0	125401	22	AAI17186	Streptomyces nous
41	68.6	7.0	1127	21	AAA02477	Human colon cancer
42	66.6	6.8	783	24	ABQ40788	Oligonucleotide fo
43	66.6	6.8	783	24	ABQ40789	Oligonucleotide fo
44	66.6	6.7	1172	24	ABQ31996	Oligonucleotide fo
45	66.6	6.7	1172	24	ABQ31997	Oligonucleotide fo

ALIGNMENTS

RESULT 1

AAV68668

ID AAV68668 standard; DNA; 979 BP.

XX

AC AAV68668;

XX

DT 03-JUN-1999 (first entry)

XX

DE Nucleotide sequence of the human Tango-81 gene.

XX

KW Human; Tango-81; host cell; recombinant protein; antibody;

KW receptor; specific binding agent; probe; primer; hybridisation;

KW amplification; mutation; genetic mapping; ss.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT CDS 58..840

FT /tag= a

FT /product= "Tango-81"

XX

PN WO9906427-A1.

XX

PD 11-FEB-1999.

XX

PF 04-AUG-1998; 98WO-US16241.

XX

PR 04-AUG-1997; 97US-0054645.

XX

PA (MILL-) MILLENNIUM BIOTHEAPEUTICS INC.

XX

PI McCarthy SA;

XX WPI: 1999-153693/13.
 DR P-PSDB; AAW84597.
 XX
 PT New nucleic acid encoding human Tango-78, -79 and -81 proteins -
 PT useful for diagnosis and treatment of Tango-associated diseases
 XX
 PS Claim 1; Fig 3; 67pp; English.
 XX
 CC This is the nucleotide sequence encoding the human Tango-81
 CC protein used in the method of the invention. Host cells containing
 CC the Tango protein are used to produce recombinant proteins for
 CC raising antibodies. It is also used in identifying specific
 CC binding agents (including cognate receptors), which can be used to
 CC determine amounts of recombinant protein in cells or
 CC therapeutically. Antibodies or other specific binding agents, are
 CC used to detect recombinant proteins and fragments of the Tango
 CC nucleotide sequence can be used as probes or primers for detecting
 CC the Tango gene, specifically mRNA, in usual hybridisation or
 CC amplification assays. These assays are used for diagnosis of
 CC diseases associated with abnormal expression of Tango proteins,
 CC e.g. detecting mutations in the Tango gene. Fragments of the Tango
 CC nucleic acid sequence are also used for genetic mapping and
 CC chromosome identification, and as antisense, ribozyme or
 CC triplex-forming therapeutics. Antibodies may also be used to generate
 CC anti-idiotypic antibodies.
 XX
 SQ Sequence 979 BP; 112 A; 383 C; 346 G; 133 T; 5 other:

Query Match 99.8%; Score 977; DB 20; Length 979;
 Best Local Similarity 100.0%; Pred. No. 8.9e-143;
 Matches 979; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCGGCAGAGCCAGCCAGTCCGCCGCGMGRGRCGCCGCTCGCTGGGGCAGCATG 60
 DB 1 GAATTCGGCAGAGCCAGCCAGTCCGCCGCGMGRGRCGCCGCTCGCTGGGGCAGCATG 60

QY 61 CGGGGTGCGCGTCTCTGGGGCGCGGGCGCGGGCGCTGGCGCTTTGGTGTCTGCTG 120
 DB 61 CGGGGTGCGCGTCTCTGGGGCGCGGGCGCGGGCGCTGGCGCTTTGGTGTCTGCTG 120

QY 121 CTGCTCGCGCTTGTTCGGCGCGCCCGCGCTCTGCGCGCGCGGTAAAGAGACCCCGC 180
 DB 121 CTGCTCGCGCTTGTTCGGCGCGCCCGCGCTCTGCGCGCGCGGTAAAGAGACCCCGC 180

QY 181 GGCCTAAGCGCAGCGTCTCGCGCTTGGCTGAGACTGCGCGCTCTCGCGCTTCCGGCG 240
 DB 181 GGCCTAAGCGCAGCGTCTCGCGCTTGGCTGAGACTGCGCGCTCTCGCGCTTCCGGCG 240

QY 241 TCAGTGGCGCGAGTGAGCGCGCGCGCGCGGTGCGAGCTGCGCGCTTCCGGCGAT 300
 DB 241 TCAGTGGCGCGAGTGAGCGCGCGCGCGGTGCGAGCTGCGCGCTTCCGGCGAT 300

QY 301 CTGCTGGAGCGCAACTCTAGAGCGCGCGCGCGCGAGGCGTGGAGATCAG 360
 DB 301 CTGCTGGAGCGCAACTCTAGAGCGCGCGCGCGAGGCGTGGAGATCAG 360

QY 361 CAGCGCGCGTCTTGGCGCGAGTCTGCGCTGTGGGGCGCGCGCGCGCAACTCTGATCCG 420
 DB 361 CAGCGCGCGTCTTGGCGCGAGTCTGCGCTGTGGGGCGCGCGCGCGCAACTCTGATCCG 420

QY 421 GCTCTGGGCTTGGAGCAGCAGCGCGCGCGCTGCGAGCTGCGCTGCGGCTCTGCTC 480
 DB 421 GCTCTGGGCTTGGAGCAGCAGCGCGCGCGCTGCGAGCTGCGCTGCGGCTCTGCTC 480

QY 481 CGCGCGCGCTTGGAGCTTGGCGCGCTAGCAGCGCGCGCTTGGCGCGCGCGCGCGCG 540
 DB 481 CGCGCGCGCTTGGAGCTTGGCGCGCTAGCAGCGCGCGCTTGGCGCGCGCGCGCGCG 540

QY 541 CGCGCGCTCGACCGCGCGCGCGCGCTAGCAGCAGCGCGCGCGCGCGCGCGCGCG 600
 DB 541 CGCGCGCTCGACCGCGCGCGCGCGCTAGCAGCAGCGCGCGCGCGCGCGCGCGCG 600

QY 601 GAGCAGCGCAGCAGACACCCGACGTCGAGCGCGCGAGCTGTTGAGTACTTGTCTGGACGG 660
 DB 601 GAGCAGCGCAGCAGACACCCGACGTCGAGCGCGCGAGCTGTTGAGTACTTGTCTGGACGG 660

QY 661 ATTCTTGGGGAAGCGGAGTCTCCGAGGGGTGGCAGCGCGCGCGCTCGGCGGTGCC 720
 DB 661 ATTCTTGGGGAAGCGGAGTCTCCGAGGGGTGGCAGCGCGCGCGCTCGGCGGTGCC 720

QY 721 GCCGACACGATGTGGGCTCTGAGCTGCCCTGAGGCGCTGCTGGGGCGCTGCTGCTG 780
 DB 721 GCCGACACGATGTGGGCTCTGAGCTGCCCTGAGGCGCTGCTGGGGCGCTGCTGCTG 780

QY 781 GTGAACGCCCTAGAGACCCCGCGCCAGTGTGTCACGCGCGCGCTTGTCCACCCCTGA 840
 DB 781 GTGAACGCCCTAGAGACCCCGCGCCAGTGTGTCACGCGCGCGCTTGTCCACCCCTGA 840

QY 841 GCATGCGCGGATCCCTGTCACCTGGGACCCAGAGTGGCCCGCGCATCCCGCCACCG 900
 DB 841 GCATGCGCGGATCCCTGTCACCTGGGACCCAGAGTGGCCCGCGCATCCCGCCACCG 900

QY 901 GACTGCTCCCGCGCAGCAGCTCCAGAGCACTTACCCCGCGCAGCGCGCTCTCACCCG 960
 DB 901 GACTGCTCCCGCGCAGCAGCTCCAGAGCACTTACCCCGCGCAGCGCGCTCTCACCCG 960

QY 961 AGGATCCCTACCCCGCTGCG 979
 DB 961 AGGATCCCTACCCCGCTGCG 979

RESULT 2
 AA158161
 ID AA158161 standard; cDNA; 1030 BP.
 XX
 AC AA158161;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 364.
 XX
 KW Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000WO-US34263.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang J, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI: 2001-442253/47.
 DR P-PSDB; AAM39005.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders

Immunosuppressant and cytostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localised neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: Immune system suppression, Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis and inflammation, leukaemias and C.N.5 disorders.

Note: The sequence data for this patent did not form part of the printed specification.

Sequence 1041 BP: 145 A: 356 C: 402 G: 138 T: 0 other: 0

Query Match

```
Query Match      98.2%; Score 961.4; DB 22; Length 1041;
Best Local Similarity 99.4%; Pred. No. 2.3e-140;
Matches 959; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
```

[illegible]

Qy	734	TGGGCTCTGAGCTGCCCCCTGAGGGCGTGCTGGGGCGCTGCTGTGCTGTGAACAGCCTAG	793
Db	727	TGGGCTCTGAGCTGCCCCCTGAGGGCGTGCTGGGGCGCTGCTGTGCTGTGAACAGCCTAG	786
Qy	794	AGACCCGGGGCCCGCAGGTGCTTGACGGCGCTCTTGCCACCTCAGCACACTGCCCGGAT	853
Db	787	AGACCCGGGGCCCGCAGGTGCTTGACGGCGCTCTTGCCACCTCAGCACACTGCCCGGAT	846
Qy	854	CCCTGTGACCCCTGGGACCCAGAAAGTGGCCCGCCGATCCGGCCACCAAGGACTGCTCCCGCG	913
Db	847	CCCGTGTGACCCCTGGGACCCAGAAAGTGGCCCGCCGATCCGGCCACCAAGGACTGCTCCCGCG	906
Qy	914	CAGCACGCTCCAGAGCAACTTACCCGGGCCAGCCAGCCCTCTCTACCCGAGGATCCCTTACCC	973
Db	907	CAGCACGCTCCAGAGCAACTTACCCGGGCCAGCCAGCCCTCTCTACCCGAGGATCCCTTACCC	966
Qy	974	CCTGGC	979
Db	967	CCTGGC	972

Matches	955;	Conservative	5;	Mismatches	2;	Indels	4;	Gaps	4;
----------------	-------------	---------------------	-----------	-------------------	-----------	---------------	-----------	-------------	-----------

RESULT 9	
ABK69773	
ID	ABK69773 standard; cDNA; 1050 BP.
XX	
XX	AC
XX	ABK69773;
XX	
XX	15-JUL-2002 (first entry)
XX	
XX	Human secreted protein gene 48.
XX	
XX	Human; ss; gene; secreted protein; gene therapy; immunosuppressive;
KW	antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
KW	vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
KW	rhumate; fungicide; ophthalmological; autoimmune disease; neoplasm;
KW	rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;
KW	cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;
KW	angiogenesis; nervous system disorder; Alzheimer's disease; infection;
KW	ocular disorder; corneal infection; wound healing; skin aging;
KW	epithelial cell proliferation; food additive.
XX	
XX	Homo sapiens.
OS	
XX	WO200226931-A2.
PN	
XX	
XX	04-APR-2002.
PD	
XX	
XX	24-SEP-2001; 2001WO-US29871.
PF	
XX	
XX	25-SEP-2000; 2000US-234925P.
PR	
PR	12-JAN-2001; 2001WO-US00911.
XX	
XX	(HUMA-) HUMAN GENOME SCI INC.
XX	
XX	Ruben SM, Komatsoulis G, Duan DR, Rosen CA, Moore PA, Shi Y;
PI	Lafleur DW, Olsen H, Brewer LA, Florence KA, Young PE, Soppet DR;
PI	Endress GA, Mucenski M, Ebner R;
XX	
XX	WPI; 2002-362489/39.
DR	P-PSDB; ABG33910, ABG34000.
DR	

PS Claim 1: page 1189; 1478pp; English.

AA The invention relates to an isolated nucleic acid molecule (or its
CC fragment, homologue complement or allelic variant) encoding a human
CC secreted protein (and its fragment, domain, epitope, variant, secreted
CC form and species variant). Also included are a recombinant vector
CC comprising the nucleic acid, a recombinant host cell comprising the

vector, an antibody against the secreted protein, a recombinant host cell that expresses the secreted protein and a method of identifying a binding partner of the secreted protein. The nucleic acid and protein are used to prevent, diagnose, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep for example autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection. Many other diseases and disorders are listed in the specification. The polypeptides can also be used to aid wound healing in the specification. The polypeptides can also be used to aid wound healing an epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities. The present sequence encodes a novel human secreted protein of the invention.

Sequence 1050 BP; 172 A; 398 C; 344 G; 136 T; 0 other;

Query Match 93.2%; Score 912.8; DB 24; Length 1050;
Best Local Similarity 98.9%; Pred. No. 7.2e-133;
Matches 955; Conservative 5; Mismatches 2; Indels 4;

Qy	14	GGCCAGCCAGTCCCGCCGSGYMCRRRRCCCGGCTCGCTGGGCGACATGCGGGGTCCGCG	73
Db	11	GGCCAGCCAGTCCCGCCGCTCCGAGCCCGGCTCGCTGGGCGACATGCGGGGTCCGCG	70
Qy	74	TGCTCTGGGGGCCCGGGCCGGGGGGCGTGGGCCCTTTTGGTGTGCTGTGCTGTGCGGCTGT	133
Db	71	TGCTCTGGGGGGCCGCGGGCCGGGGGGCGTGGGCCCTTTTGGTGTGCTGTGCTGTGCGGCTGT	130
Qy	134	TTGGGCGCCGCCCGCGCTCTGCGCGCGGCGGTAAAGGAGCCCCGCGGCTAAAGCGAG	193
Db	131	TTGGGCGCCGCCCGCGCTCTGCGCGCGGCGGTAAAGGAGCCCCGCGGCTAAAGCGAG	190
Qy	194	CGTCTCCGCCCTTGGCTGAGACTGGGCGCTCTCCGCGCTTCCGGCGGTGAGTGCCTCCGAG	253
Db	191	CGTCTCCGCCCTTGGCT - AGACTGGGCTCTCTCGCGCTTCCGGCGGTGAGTGCCTCCGAG	249
Qy	254	GTGAGGCGCGGGGGCGGTGCAGAGCTGGCGCGGCGCTGGCGCATCTGCTGGAGCGCG	313
Db	250	GTGAGGCGCGGGGGCGGTGCA - GAGCTGGCGCGGCGCTGGCGCATCTGCTGGGA - GCG	307
Qy	314	AAGCTCAGGAGCGGGCGGCGAGCGCAGAGAGCTGAGGATCAGCAGCGCGCGGTCC	373
Db	308	AAGCTCAGGAGCGGGCGGCGGCGAGGCGCA - GAGGCTGAGGATCAGCAGCGCGGTCC	366
Qy	374	TGGCGCAGCTGCTCGCGCTGTGGGGCGCCCCCGCAACTCTGATCGGCTCTGGGCTGG	433
Db	367	TGGCGCAGCTGCTCGCGCTGTGGGGCGCCCCCGCAACTCTGATCGGCTCTGGGCTGG	426
Qy	434	ACGACGACCCGACGGCGCTTGCAGCGAGCTGGCTCGCGCTCTGCTCCGCGCCCGCTTG	493
Db	427	ACGACGACCCGACGGCGCTTGCAGCGAGCTGGCTCGCGCTCTGCTCCGCGCCCGCTTG	486
Qy	494	ACCTTCCGCGCTTAGCAGCCAGCTTGTCCCGCGCGCTCCCGCGCGGGCGCTCCGAC	553
Db	487	ACCTTCCGCGCTTAGCAGCCAGCTTGTCCCGCGCGCTCCCGCGCGGGCGCTCCGAC	546
Qy	554	CCGGCCCCGGGTACGACGACGGCCCCCGGGCGCGGATGCTGAGGAGGACGAGCGAG	613
Db	547	CCGGCCCCGGGTACGACGACGGCCCCCGGGCGCGGATGCTGAGGAGGACGAGCGAG	606
Qy	614	AGACACCCGAGCTGGACCCCGAGCTTTGAGGTACTTGTGGGAGGATTTCTTGGGGAA	673
Db	607	AGACACCCGAGCTGGACCCCGAGCTTTGAGGTACTTGTGGGAGGATTTCTTGGGGAA	666
Qy	674	GCGGGACTCCGAGGGGTGGCAGCCCCCGCGCGCTCCGCGGTGCGCCGACACGATG	733
Db	667	GCGGGACTCCGAGGGGTGGCAGCCCCCGCGCGCTCCGCGGTGCGCCGACACGATG	726

CC producing (I) in large quantities. Cells containing (II) are used for
CC the detection of ligands or receptors corresponding to membrane or
CC secretory proteins and to screen small molecule novel pharmaceuticals.
CC Antibodies directed to (I) can be used for the detection, quantification
CC and purification of (I). Activities of (I) may include cytokine and cell
CC proliferation/differentiation function, immune stimulating or suppressing
CC activity, haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity and anti-inflammatory
CC activity. (I) and (II) can be used to treat autoimmune disorders e.g.
CC multiple sclerosis, HIV infections, anaemia, burns, ulcers, osteoporosis,
CC inflammatory bowel disease and tumours. (I) and (II) can also be used for
CC wound healing, as nutritional sources or supplements e.g. as amino acid,
CC carbon or nitrogen source, to affect metabolism, catabolism, anabolism,
CC processing and utilisation of dietary fat, protein, carbohydrate,
CC vitamins and minerals, to effect behavioural characteristics, to affect
CC appetite, and can act as antigens in vaccines to raise an immune response
CC to the protein or another material cross-reactive with the protein.
XX

SQ Sequence 780 BP; 79 A; 297 C; 294 G; 110 T; 0 other;

Query Match 79.3%; Score 776.8; DB 22; Length 780;

Best Local Similarity 99.7%; Pred. No. 7.5e-112;

Matches 778; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 58 ATGGCGGGTCTGGCTCTCTGGGGCGCGGGCGGGGGGGCTTTTGGTGTCTG 117
DB |||||
QY 118 CTGCTGCTGGCGCTGTTTGGCGCGCCCGCGCTCTGGCGCGCGGTAAGAGAGCCC 177
DB |||||
QY 61 CTGCTGCTGGCGCTGTTTGGCGCGCCCGCGCTCTGGCGCGCGGTAAGAGAGCCC 120
DB |||||
QY 178 CGCGGGCTAAGCGACAGCTCTCCGCCCTTGGCTGAGACTGGCGCTCTCGCCCTTCCGG 237
DB |||||
QY 121 CGCGGGCTAAGCGACAGCTCTCCGCCCTTGGCTGAGACTGGCGCTCTCGCCCTTCCGG 180
DB |||||
QY 238 CGGTGAGTCCCGAGGTGAGCGGGCGGGGGCGGTGCGAGGCTGGCGGGCGCTGGCG 297
DB |||||
QY 181 CGGTGAGTCCCGAGGTGAGCGGGCGGGGGCGGTGCGAGGCTGGCGGGCGCTGGCG 240
DB |||||
QY 298 CATCTGCTGAGGCGCAACCTCAGGAGCGGGCGGGCGCGAGCGCAGGAGGCTGAGGAT 357
DB |||||
QY 241 CATCTGCTGAGGCGCAACCTCAGGAGCGGGCGGGCGCGAGCGCAGGAGGCTGAGGAT 300
DB |||||
QY 358 CAGCAGCGCGCTCTGGCGAGCTGCTGCGCGTCTGGGGCGCCCGCCCAACTCTGAT 417
DB |||||
QY 301 CAGCAGCGCGCTCTGGCGAGCTGCTGCGCGTCTGGGGCGCCCGCCCAACTCTGAT 360
DB |||||
QY 418 CCGGCTCTGGGCTTGGACGACGACCGCGCGCGCTGCGAGCGAGCTGCTCGGCTCTG 477
DB |||||
QY 361 CCGGCTCTGGGCTTGGACGACGACCGCGCGCGCTGCGAGCGAGCTGCTCGGCTCTG 420
DB |||||
QY 478 CTCGGCGCGCTTGAACCTGCGCCCTAGCAGCCAGCTTCTCCCGCGCGCGCTCCCG 537
DB |||||
QY 421 CTCGGCGCGCTTGAACCTGCGCCCTAGCAGCCAGCTTCTCCCGCGCGCGCTCCCG 480
DB |||||
QY 538 GCGCGCGCTCGACCCCGCGCGCTTACGACGAGCGCGCGCGCGCGCGGATGCT 597
DB |||||
QY 481 GCGCGCGCTCGACCCCGCGCGCTTACGACGAGCGCGCGCGCGCGCGGATGCT 540
DB |||||
QY 598 GAGGAGCGGCGACGACACCGGAGCTGGAGCCCGAGCTGTGAGGTACTTGTGGGA 657
DB |||||
QY 541 GAGGAGCGGCGACGACACCGGAGCTGGAGCCCGAGCTGTGAGGTACTTGTGGGA 600
DB |||||
QY 658 CGGATTTCTCGGGAAGCGGAGTCCGAGGGGGTGGAGCCCGCGCGCGCTCCCGCT 717
DB |||||
QY 601 CGGATTTCTCGGGAAGCGGAGTCCGAGGGGGTGGAGCCCGCGCGCGCTCCCGCT 660
DB |||||
QY 718 GCCCGCGACACGATGTGGCTCTGAGCTGCCCGCTGAGGGCGTGTGGGGGGCTGCTG 777
DB |||||
QY 661 GCCCGCGACACGATGTGGCTCTGAGCTGCCCGCTGAGGGCGTGTGGGGGGCTGCTG 720
DB |||||

QY 778 CGTGTGAACGCTAGAGACCCCGCGCGCGCGCGCTGCTGACGCGGCTCTTGGCACCC 837
DB |||||
QY 721 CGTGTGAACGCTAGAGACCCCGCGCGCGCGCGCTGCTGACGCGGCTCTTGGCACCC 780
DB |||||

RESULT 11

AAI58162
ID AAI58162 standard; cDNA; 919 BP.

XX AAI58162;
AC AAI58162;

DT 22-OCT-2001 (first entry)

XX Human polynucleotide SEQ ID NO 365.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.

XX Homo sapiens.
OS

XX WO200153312-A1.
PN

XX 26-JUL-2001.
PD

XX 26-DEC-2000; 2000WO-US34263.
PF

XX 21-JAN-2000; 2000US-0488725.
PR

XX 25-APR-2000; 2000US-0552317.
PR

XX 09-JUL-2000; 2000US-0598042.
PR

XX 19-JUL-2000; 2000US-0620312.
PR

XX 03-AUG-2000; 2000US-0653450.
PR

XX 14-SEP-2000; 2000US-0662191.
PR

XX 19-OCT-2000; 2000US-0693036.
PR

XX 29-NOV-2000; 2000US-0727344.
PR

XX (HYSE-) HYSEQ INC.
PA

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI; 2001-442253/47.
DR

XX P-PSDB; AAM39006.
DR

XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -

XX Claim 1; SEQ ID NO 365; 10078pp; English.
PS

XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemia and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.

XX Sequence 919 BP; 122 A; 365 C; 308 G; 124 T; 0 other;
SQ

Query Match 74.5%; Score 729.4; DB 22; Length 919;


```

XX  SQ  Sequence 991 BP; 158 A; 365 C; 315 G; 153 T; 0 other;
Query Match      57.8%; Score 565.4; DB 21; Length 991;
Best Local Similarity 78.1%; Pred. No. 3.3e-79;
Matches 698; Conservative 5; Mismatches 184; Indels 7; Gaps 2;

QY  15  GCACGACGTCGCCGCGMCRGCGCGTCCGTCGGGGAGCATGCGGGGTCGCGCGT 74
Db  16  GACAGCAGTCCCGCGTCCGAGACCGCGTCTGTGGGAGCATGCGGGGTCGCGCGT 75
QY  75  GCTCTGGGGCGCGGGCGCGGCGGTCGCGCGCTTTTGGTCTGCTGCTCGGCCCTGTT 134
Db  76  GCTCTGCGGGCGCGGGCGGGCGGCTCGCGCATTTTGGTCTGCTGCTTGGGCGCTTCT 135
QY  135  TCGGCGCGCGCGCGCGTCTGCGCGCGCGGTAAGAGCGCGCGCGCGTAAAGCGCAGC 194
Db  136  GAGGCTGCCCGCGCGCGTCTGAGGAGGCGCGTGAAGGAGCGCGCGAGTCTGAGCGCAGC 195
QY  195  GTCTCCGCGCTTGGCTGAGACTGGCGTCTCGCGCGTTCGCGGGTCACTGCGCCGAGG 254
Db  196  ATCCGCGCGCTTGGTTGAGACGACACTCCCGTCCGCTTCCGCTGCGGCGCGTCCCGGAGG 255
QY  255  TGAGCGCGCGCGCGTGCAGGAGTGGCGCGGCGCTGCGCATCTGCTGGAGGCGCA 314
Db  256  AGAGCGCGCGTGGTGCAGAGTGGCGCGGCGCGCGTGGCGCAACCGTGTGGAGGCCGA 315
QY  315  ACCTGAGGAGCGCGCGCGCGCGAGGCGAGGCTGAGGATCAGCAGCGCGCGCTTCCT 374
Db  316  GAGACAGGAGCGCGCGCGTCTGAGGCGCAGGAGGCTGAGGATCAGCAGCGCGTGTCTCT 375
QY  375  GCGCGAGCTCTGCGCGTCTGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 434
Db  376  GCGCGAGCTCTGCGCGCTTGGGGCTCTCCGCGTCTCGCGCGCGCGCGCGCGCGCGCG 435
QY  435  CGAGCAGCGCGCGCGCTGACGAGCTGCTGCGCGCTGCTGCGCGCGCGCGCGCGCGCTGA 494
Db  436  CGATGACCGCGAGCTCCAGCTGCACAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGTGA 495
QY  495  CCTGCGCGCGCTAGCAGCGCGCTTGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 554
Db  496  CCGCGCGCGCGTGCAGCGCGCAACTGTGCCCGC-----CCCTGCGCGCTCGCGCGCGCG 549
QY  555  CCGCGCGCGCGTCTACGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 614
Db  550  CCGCGCGCGCGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 609
QY  615  GACACCGAGCTGGACCGCGAGCTGTGAGGTACTTGTGCGGACGAGTCTTGGGGAAG 674
Db  610  GACTCTGAGCTGAGCGCTGAGCTGAGTACTGAGTACTGAGTACTGAGTACTGAGTACTGAG 669
QY  675  CGCGGACTCCGAGGGGTGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGATG 734
Db  670  TTCGAGCCAGAGGCTGCTCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGATTT 729
QY  735  GGGCTCTGAGCTGCCCGCTGAGGCGTGTGCGGGCGCGCTGCTGCTGTAAGCGCGCTAGA 794
Db  730  GGTTCGCGAGGCGCGCGCGTCTGAGACGCTACTGCGGGGCTCTGTATACGCTCAAGCGCTGA 789
QY  795  GACCGCGCGCGCGCGAGTGTGCTGACGCGCGCTTGTGCCACCTGAGCACTGCCCGGATC 854
Db  790  GAACCGCTGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGATC 848
QY  855  CCGTGTGACCGTGGGACCGCGAGGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 908
Db  849  CTGACGCGCGCTGGAGCCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAC 902
```

RESULT 13
ABK69876
ID ABK69876 standard; DNA; 4503 BP.
XX
AC ABK69876;

```

XX  DT  15-JUL-2002 (first entry)
XX  DE  Human secreted protein gene 49 genomic DNA fragment #2.
XX  KW  Human; ds; secreted protein; gene therapy; immunosuppressive;
KW  anarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
KW  vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
KW  virucide; fungicide; opthalmological; autoimmune disease; neoplasm;
KW  rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;
KW  cardiovascular disorder; cerebrovascular disorder; cerebral ischemia;
KW  angiogenesis; nervous system disorder; Alzheimer's disease; infection;
KW  ocular disorder; corneal infection; wound healing; skin aging;
KW  epithelial cell proliferation; food additive.
XX  OS  Homo sapiens.
XX  PN  WO200226931-A2.
XX  PD  04-APR-2002.
XX  PF  24-SEP-2001; 2001WO-US29871.
XX  PR  25-SEP-2000; 2000US-234925P.
XX  PR  12-JAN-2001; 2001WO-US00911.
XX  PA  (HUMA-) HUMAN GENOME SCI INC.
XX  PI  Ruben SM, Komatsoulis G, Duan DR, Rosen CA, Moore PA, Shi Y;
PI  Lafleur DW, Olsen H, Brewer LA, Florence KA, Young PE, Soppet DR;
PI  Endress GA, Mucenski M, Ebner R;
XX  WI  2002-362489/39.
XX  PT  Novel 71 isolated secreted polypeptides and polynucleotides encoding
PT  the polypeptides, useful for treating Huntington's disease, sepsis,
PT  meningitis, thrombocytopenia, haemolytic anaemia, rheumatoid arthritis,
PT  asthma
XX  Example 2: Page 1404-1406; 1478pp; English.
XX  CC  The invention relates to an isolated nucleic acid molecule (or its
CC  fragment, homologue complement or allelic variant) encoding a human
CC  secreted protein (and its fragment, domain, epitope, variant, secreted
CC  form and species variant). Also included are a recombinant vector
CC  comprising the nucleic acid, a recombinant host cell comprising the
CC  vector, an antibody against the secreted protein, a recombinant host cell
CC  that expresses the secreted protein and a method of identifying a binding
CC  partner of the secreted protein. The nucleic acid and protein are used to
CC  prevent, diagnose, treat or ameliorate a medical condition in e.g.
CC  humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep
CC  for example autoimmune diseases e.g. rheumatoid arthritis,
CC  hyperproliferative disorders e.g. cardiac arrest, cerebrovascular disorders
CC  e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC  Alzheimer's disease, infections caused by bacteria, viruses and fungi and
CC  ocular disorders e.g. corneal infection. Many other diseases and
CC  disorders are listed in the specification. The polypeptides can also be
CC  used to aid wound healing an epithelial cell proliferation, to prevent
CC  skin aging due to sunburn, to maintain organs before transplantation, for
CC  supporting cell culture of primary tissues, to regenerate tissues and in
CC  chemotaxis. The polypeptides can also be used as a food additive or
CC  preservative. The polypeptides can also be used as storage capabilities. The present
CC  sequence represents a ds DNA fragment of the gene for a novel human
CC  secreted protein of the invention.
XX  SQ  Sequence 4503 BP; 783 A; 1181 C; 1314 G; 1225 T; 0 other;

Query Match      48.4%; Score 474.2; DB 24; Length 4503;
Best Local Similarity 99.4%; Pred. No. 3.3e-65;
Matches 476; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY  170  AGGAGCGCGCGCGCTAAGCGCAGCGTCTCCGCCCTTGGCTGAGACTGCGCGCTCTCGGC
```

```
Db 3254 AGAGCCCGCCGCTAAGCAGCGTCTCCGCCCTTGCTGAGACTGCGCTCTCTCGCC 3313
QY 230 GCTTCCGGCGGTACGTGCCCCAGGTGAGCGCGCGGGCGGTGAGGAGCTGGCGCGGG 289
Db 3314 GCTTCCGGCGGTACGTGCCCCAGGTGAGCGCGCGGGCGGTGAGGAGCTGGCGCGGG 3373
QY 290 CCGTGGCGCATCTGCTGGAGGCCGAACTGAGAGCGCGCGGGCGGCGGAGGAGG 349
Db 3374 CGCTGGCGCATCTGCTGGAGGCCGAACTGAGAGCGCGCGGGCGGCGGAGGAGG 3433
QY 350 CTGAGATCAGAGCGCGGCTCTGCGCAGCTGCTGGCGCTCTGGCGCGCGCGCGCA 409
Db 3434 CTGAGATCAGAGCGCGGCTCTGCGCAGCTGCTGGCGCTCTGGCGCGCGCGCGCA 3493
QY 410 ACTGATCCGCGCTCTGGGCTTGGAGCAGCAGCCCGAGCGGCTGCGAGCGAGCTCGTC 469
Db 3494 ACTGATCCGCGCTCTGGGCTTGGAGCAGCAGCCCGAGCGGCTGCGAGCGAGCTCGTC 3553
QY 470 GCGCTCTGCTCGCGCGCGCTTGACCTGCGCGCTAGCAGCCAGCTGTCTCCCGCGC 529
Db 3554 GCGCTCTGCTCGCGCGCGCTTGACCTGCGCGCTAGCAGCCAGCTGTCTCCCGCGC 3613
QY 530 CGGTCCCGCGCGCGCTCGCAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 589
Db 3614 CGGTCCCGCGCGCGCTCGCAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3673
QY 590 CGGATGCTGAGGAGCAGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 648
Db 3674 CGGATGCTGAGGAGCAGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 3732

RESULT 14
ABQ31154/c
ID ABQ31154 standard; DNA; 652 BP.
XX
AC ABQ31154;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 17745.
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
XX SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
XX
PS 05-SEP-2000; 2000DE-1044543.
XX
FA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA
XX
XX Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
XX This invention describes a novel method for determining the degree of
XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a
XX genomic sample of DNA. The sample is treated chemically to convert
```

```
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNPs); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 652 BP; 103 A; 81 C; 233 G; 235 T; 0 other;
```

```
Query Match 30.8%; Score 301.6; DB 24; Length 652;
Best Local Similarity 77.1%; Pred. No. 1.9e-38;
Matches 367; Conservative 0; Mismatches 109; Indels 0; Gaps 0;
QY 173 AGCCCGCGCGCTAAGCGCAGCGTCTCCGCCCTTGCTGAGACTGCGCTCTCTCGCGCT 232
Db 511 AACC CGCGCGCTAAGCGCAGCGTCTCCGCCCTTAACTAAACTAACGCTCTCTCGCGCT 452
QY 233 TCCGCGCGCTGAGTCCCGCGGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 292
Db 451 TCCGCGCGCTGAGTCCCGCGGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 392
QY 293 TGGCGCATCTGCTGGAGCGCGAAGCTCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCT 352
Db 391 TAACGCATCTACTTAAACCGGAGCTCAAAACGAGACGCGACCGCAACGCAAACTA 332
QY 353 AGGATCAGCAGCGCGCGCTCTCTGCGCAGCTGCTGCGCGCTGCGCGCGCGCGCGCG 412
Db 331 AAATCAACAAACGCGCGCTCTTAAACGCACTACTACGCGCTTAAACGCGCGCGCGCACT 272
QY 413 CTGATCCGCGCTGCGGCTTGGAGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCT 472
Db 271 CTATCCGACCTTAAACCTAAACGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCT 212
QY 473 CTCTGCTCCGCGCGCGCTTGAACCTTGCCTGCGCGCTAGCAGCGCGCGCGCGCGCGCG 532
Db 211 CTCTACTCCGCGCGCGCTTAAACCTTACCGCTACCGCGCTCGCAACCACTTATCCCGCGCG 152
QY 533 TCCCGCGCGCGCGCTCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 592
Db 151 TCCCGCGCGCGCGCTCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 92
QY 593 ATGCTGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 648
Db 91 ATACTAAAAAACAACGCAACCGGAGCTAAACCGCGCGCGCGCGCGCGCGCGCGCG 36
```

RESULT 15
ABQ31155

ID ABQ31155 standard; DNA; 652 BP.

XX ABQ31155;

AC ABQ31155;

DT 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 17746.

XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.

XX OS Homo sapiens.

XX WO200218632-A2.
XX PD 07-MAR-2002.
XX PF 01-SEP-2001; 2001WO-EP10074.
XX PR 01-SEP-2000; 2000DE-1043826.
XX PR 05-SEP-2000; 2000DE-104543.
XX PA (EPIG-) EPIGENOMICS AG.
XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX DR WPI; 2002-371829/40.
XX
XX Determining the degree of cytosine methylation in genomic DNA, useful
XX for diagnosis and prognosis, comprises selective hybridization of
XX amplicons from chemically treated DNA -
XX
XX Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
XX This invention describes a novel method for determining the degree of
XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a
XX genomic sample of DNA. The sample is treated chemically to convert
XX cytosine (C) but not methylated C, to uracil, then part of the genomic
XX DNA that contains the target C is amplified to form a labeled amplicon.
XX The amplicon is hybridised to two classes, each with at least one
XX member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
XX and the degree of hybridisation to both classes is determined from the
XX label on the amplicon. From the ratio of labels hybridised to the two
XX classes of oligomers, the degree of methylation is calculated. The method
XX is used: (i) for diagnosis and/or prognosis of side effects of
XX therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
XX of the central nervous, cardiovascular, gastrointestinal and respiratory
XX systems etc., particularly by detecting mutations or single nucleotide
XX polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
XX types and for investigating cell differentiation. The method allows the
XX methylation status of many C residues to be determined simultaneously.
XX ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
XX method for determining the degree of cytosine methylation described in
XX the disclosure of the invention.
XX
XX Sequence 652 BP; 235 A; 233 C; 81 G; 103 T; 0 other;

QY 533 TCCCGCGCGGCGCTCGGACCCCGCGCGCGCTTACGACGAGCGCGCGCGCGCGG 592
Db 502 TCCCGCGCGGAGCGCTCCGACCCCGCGCGCGGATCTACGACGAGCGCGCGCGCGG 561
QY 593 ATGCTGAGGAGCGCAGCGAGGACACCCGACGTTGGACCCCGGAGCTGTGAGGTAC 648
Db 562 ATACTAAACAAACAGCAGGAAACCCGACGTAACCCCGAACTATTAAATAC 617

Search completed: April 12, 2003, 17:22:03
Job time : 292 secs

Query Match 30.8%; Score 301.6; DB 24; Length 652;
Best Local Similarity 77.1%; Pred. No. 1.9e-38;
Matches 367; Conservative 0; Mismatches 109; Indels 0; Gaps 0;
QY 173 AGCCCCGGCGCTAAGCGCAGCGTCTCGCCCTTGGCTGAGACTGGCGCTCCTCGCGGCT 232
Db 142 AACCCCGGAGCTTAACGCAACGCTCTCGCCCTTAACTAAACTTAACGCTCTCTCGCGCT 201
QY 233 TCCGGCGGTAGTCCCGGAGTGAGCGGGGGGGTGCAGGAGTGGCGGGGCGC 292
Db 202 TCCGACGATCAATACCCCGAATAAAGAGGAAACGATACAAAACATACGCGAACGC 261
QY 293 TGGCGCATCTGCTGGAGCGCGCAAGCTCAGGAGCGGGCGCGGCGGAGCGAGGAGCTG 352
Db 262 TAACGCATCTACTAAAAACCGACGCTCAAAAACGACGCAACCGAACAACCACTA 321
QY 353 AGGATCAGCAGGCGCGGCTCTGGCGGAGCTGTGGCGGCGCGCGCGCAACT 412
Db 322 AAAATCAACAACCGCGCTCTTACGCAACTACTACGCGCTTAAACGCGCGCGCAACT 381
QY 413 CTGATCCGGCTCTGGGCTTGACGACGACCGCGCGCTGCGCGAGCTGCTCGCG 472
Db 382 CTATCCGACTTAAACCTAAACGACGACCGCGCGCGCTTACACGCACTGCTCGCG 441
QY 473 CTCGTCTCCGCGCGCGCTTGACCTTCCGCGCTTACGAGCCAGCTTGTCCCGCGCGCG 532
Db 442 CTCCTACTCCGCGCGCGCTTAAACCTTACGCGCTTACGCGCTTACCGCACTTATCCCGCGCGCG 501

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2003, 14:08:39 ; Search time 37 Seconds
(without alignments)
936.356 Million cell updates/sec

Title: US-09-803-589-6
Perfect score: 1319
Sequence: 1 MAGSPLWPGAGGVGLLV.....RVKRLTPAPQVPRRLPP 260

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1319	100.0	260	20	AAW84597
2	1319	100.0	260	21	AAY71959
3	1319	100.0	260	22	AAW39005
4	1319	100.0	260	22	AAW88596
5	1105.5	83.8	223	22	AAW39006
6	1019	77.3	258	21	AAY71960
7	1015	77.0	262	21	AAY71962
8	308	23.4	148	22	AAE06088
9	308	23.4	148	23	ABG33910
10	308	23.4	149	21	AAY87111

11	308	23.4	166	21	AAY87199	Human secreted pro
12	308	23.4	166	22	ABG06176	Human gene 48 enco
13	308	23.4	166	23	ABG34000	Human secreted pro
14	259	19.6	69	21	AAY71961	Rat pituitary horm
15	136.5	10.3	388	23	ABG60300	Lymphoma associate
16	136.5	10.3	388	23	ABG09271	G protein-coupled
17	136.5	10.3	451	22	AAU68528	Human novel cytol
18	134.5	10.2	317	22	AAU40791	Human polypeptide
19	134.5	10.2	317	22	AAU40792	Human polypeptide
20	124	9.4	720	23	AAE20839	Human gene 7 encod
21	124	9.4	766	23	AAE20838	Human gene 7 encod
22	124	9.4	973	23	ABG64875	Human albumin fusi
23	124	9.4	973	23	ABG64875	Human gene 7 encod
24	123.5	9.4	597	23	ABG20799	Human secretory po
25	123	9.3	251	22	AAU04906	Micromonospora eve
26	121.5	9.2	481	22	ABG05355	Novel human diago
27	119.5	9.1	248	15	AAE46912	GGF segment E. Ho
28	119.5	9.1	248	15	AAE55853	GGF segment E. Ho
29	119.5	9.1	248	17	AAW09358	Human glial growth
30	119.5	9.1	248	17	AAW09371	Human glial growth
31	119.5	9.1	248	17	AAW09372	Human glial growth
32	119.5	9.1	248	20	AAE87445	Human neurogulin g
33	119.5	9.1	422	15	AAE55854	GGF-II encoded by
34	119.5	9.1	422	15	AAE46923	GGF-II encoded by
35	119.5	9.1	422	16	AAE67258	Human glial cell g
36	119.5	9.1	422	17	AAW09371	Human neurogulin g
37	119.5	9.1	422	17	AAW09372	Human GGF2. Homo
38	119.5	9.1	422	17	AAE96081	Glial growth facto
39	119.5	9.1	422	17	AAE87466	Glial growth facto
40	119.5	9.1	422	17	AAE87467	Glial growth facto
41	119.5	9.1	422	17	AAE86628	Mature hGGF2. Hom
42	119.5	9.1	422	22	AAE67901	Human neurogulin g
43	119.5	9.1	422	22	AAE67939	Human neurogulin g
44	119	9.0	404	19	AAE68408	Aujesky's disease
45	119	9.0	421	22	ABE71817	Drosophila melanog

ALIGNMENTS

RESULT 1
AAW84597
ID AAW84597 standard; Protein; 260 AA.
XX AAW84597;
XX AC
XX DT 03-JUN-1999 (first entry)
XX DE Amino acid sequence of the human Tango-81 protein.
XX Human; Tango-81; host cell; recombinant protein; antibody;
KW receptor; specific binding agent; probe; primer; hybridisation;
KW amplification; mutation; genetic mapping.
XX OS Homo sapiens.
XX WO906427-A1.
XX PD 11-FEB-1999.
XX PF 04-AUG-1998; 98WO-US16241.
XX PR 04-AUG-1997; 97US-0054645.
XX (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
XX PI McCarthy SA;
XX WPI; 1999-153693/13.
XX N-PSDB; AAV68668.
XX New nucleic acid encoding human Tango-78, -79 and -81 proteins -
XX useful for diagnosis and treatment of Tango-associated diseases
PT

XX Claim 1; Fig 3; 67pp; English.

XX This is the amino acid sequence of the human Tango-81 protein

CC used in the method of the invention. Host cells containing

CC the Tango protein are used to produce recombinant proteins for

CC raising antibodies. It is also used in identifying specific

CC binding agents (including cognate receptors), which can be used to

CC determine amounts of recombinant protein in cells or

CC therapeutically. Antibodies or other specific binding agents, are

CC used to detect recombinant proteins and fragments of the Tango

CC nucleotide sequence can be used as probes or primers for detecting

CC the Tango gene, specifically mRNA, in usual hybridisation or

CC amplification assays. These assays are used for diagnosis of

CC diseases associated with abnormal expression of Tango proteins,

CC e.g. detecting mutations in the Tango gene. Fragments of the Tango

CC nucleic acid sequence are also used for genetic mapping and

CC chromosome identification, and as antisense, ribozyme or

CC triplex-forming therapeutics. Antibodies may also be used to generate

CC anti-idiotypic antibodies.

XX Sequence 260 AA;

XX Query Match 100.0%; Score 1319; DB 20; Length 260;

XX Best Local Similarity 100.0%; Pred. No. 2.9e-104;

XX Matches 260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGSPLLMGPRAGGVLVLLGLFRPPPPALCARPVKEPRGLSAASPPPLAETGAPRRFR 60

DB 1 MAGSPLLMGPRAGGVLVLLGLFRPPPPALCARPVKEPRGLSAASPPPLAETGAPRRFR 60

QY 61 RSVPRGEAAGAVQELARALAHLEAEQERARAEAEQEDQARVLAQLLRVWGAPRNSD 120

DB 61 RSVPRGEAAGAVQELARALAHLEAEQERARAEAEQEDQARVLAQLLRVWGAPRNSD 120

QY 121 PALGLDDDDPDAPAAQARALLRARLDPAALAAQLVPAPVPAALRPPPVYDDGAPGPD 180

DB 121 PALGLDDDDPDAPAAQARALLRARLDPAALAAQLVPAPVPAALRPPPVYDDGAPGPD 180

QY 181 EAGDETDPDPELLRYLLGRILASADSEGVAAAPRRRLRAADHDVSGELPPEGVLGALL 240

DB 181 EAGDETDPDPELLRYLLGRILASADSEGVAAAPRRRLRAADHDVSGELPPEGVLGALL 240

QY 241 RVKRLTAPQVPARRLLPP 260

DB 241 RVKRLTAPQVPARRLLPP 260

RESULT 2

AA771959

ID AA771959 standard; Protein; 260 AA.

AC AA771959;

XX 26-MAR-2001 (first entry)

XX Human pituitary hormone, pituitrone.

XX Human; pituitary; pituitrone; therapy; immune disorder; anaemia;

KW Digeorge syndrome; haematopoietic cell; ataxia telangiectasia;

KW Wiskott-Aldrich disorder; blood coagulation disorder; Addison's disease;

KW autoimmune disorder; multiple sclerosis; systemic lupus erythematosus;

KW SLE; hyperproliferative disorder; gene therapy; neoplasm;

KW infectious disease; immunomodulatory; cytostatic; antimicrobial.

OS Homo sapiens.

XX Key Location/Qualifiers

XX 1..33 /label= Signal_peptide

FT Peptide

FT 34..260 /label= Mature_human_pituitrone

FT Protein

FT 36..44

FT Region

FT Region /note= "Highly immunogenic"

FT 55..68

FT /note= "Highly immunogenic"

FT 116..121

FT /note= "Highly immunogenic"

FT 123..131

FT /note= "Highly immunogenic"

FT 164..182

FT /note= "Highly immunogenic"

FT 185..193

FT /note= "Highly immunogenic"

FT 206..211

FT /note= "Highly immunogenic"

FT 225..233

FT /note= "Highly immunogenic"

FT 253..258

FT /note= "Highly immunogenic"

XX WO200066778-A1.

XX 09-NOV-2000.

XX 27-APR-2000; 2000WO-US11211.

XX 30-APR-1999; 99US-0131966.

XX (HUMA-) HUMAN GENOME SCI INC.

PA Ruben SM, NI J;

PI WPI; 2000-687547/67.

DR N-PSDB; AAD02075.

DR Nucleic acids encoding a novel pituitary hormone (pituitrone), useful

PT for diagnosing, preventing and treating e.g. immune disorders,

PT hyperproliferative disorders and blood coagulation disorders -

XX Claim 11; Fig 1; 277pp; English.

XX The present sequence is a human pituitary hormone, pituitrone (clone

CC HKGL36). Pituitrone is highly expressed in pituitary

CC gland and is also expressed in brain tissues, spinal cord and kidney.

CC It may be used as antigens in the production of antibodies against

CC pituitrone and in assays to identify modulators. Pituitrone cDNAs are

CC also useful in gene therapy. Pituitrone may be used in the prevention,

CC treatment and diagnosis of diseases associated with inappropriate

CC pituitrone expression. It may be useful in treating disorders related

CC to reproductive and renal system, immune disorders, disorders of

CC haematopoietic cells (e.g., anaemia, Digeorge syndrome, ataxia

CC telangiectasia and Wiskott-Aldrich disorder), blood coagulation

CC disorders, autoimmune disorders (e.g. Addison's disease, multiple

CC sclerosis and systemic lupus erythematosus (SLE)), hyperproliferative

CC disorders (e.g. neoplasms of bone, liver and pancreas) and infectious

CC diseases.

XX Sequence 260 AA;

XX Query Match 100.0%; Score 1319; DB 21; Length 260;

XX Best Local Similarity 100.0%; Pred. No. 2.9e-104;

XX Matches 260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGSPLLMGPRAGGVLVLLGLFRPPPPALCARPVKEPRGLSAASPPPLAETGAPRRFR 60

DB 1 MAGSPLLMGPRAGGVLVLLGLFRPPPPALCARPVKEPRGLSAASPPPLAETGAPRRFR 60

QY 61 RSVPRGEAAGAVQELARALAHLEAEQERARAEAEQEDQARVLAQLLRVWGAPRNSD 120

DB 61 RSVPRGEAAGAVQELARALAHLEAEQERARAEAEQEDQARVLAQLLRVWGAPRNSD 120

QY 121 PALGLDDDDPDAPAAQARALLRARLDPAALAAQLVPAPVPAALRPPPVYDDGAPGPD 180

DB 121 PALGLDDDDPDAPAAQARALLRARLDPAALAAQLVPAPVPAALRPPPVYDDGAPGPD 180

QY 181 EEAGDETPDVPPELLRYLLGRILAGSADSEGVAAAPRRLRAADHDVGSSELPPEGVLGALL 240
Db 181 EEAGDETPDVPPELLRYLLGRILAGSADSEGVAAAPRRLRAADHDVGSSELPPEGVLGALL 240
QY 241 RVKLETPAPQVPARRLLPP 260
Db 241 RVKLETPAPQVPARRLLPP 260

RESULT 3

AAM39005

ID AAM39005 standard; Protein; 260 AA.

XX AC AAM39005;

XX DT 22-OCT-2001 (first entry)

XX DE Human polypeptide SEQ ID NO 2150.

XX KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;

XX KW peripheral nervous system; neuropathy; central nervous system; CNS;

XX KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;

XX KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

XX KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;

XX KW leukaemia.

XX OS Homo sapiens.

XX PN WO200153312-A1.

XX PD 26-JUL-2001.

XX PF 26-DEC-2000; 2000WO-US34263.

XX PR 21-JAN-2000; 2000US-0488725.

XX PR 25-APR-2000; 2000US-0552317.

XX PR 09-JUL-2000; 2000US-0598042.

XX PR 19-JUL-2000; 2000US-0620312.

XX PR 03-AUG-2000; 2000US-0653450.

XX PR 14-SEP-2000; 2000US-0662191.

XX PR 19-OCT-2000; 2000US-0693036.

XX PR 29-NOV-2000; 2000US-0727344.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

XX PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

XX PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX DR WPI; 2001-442253/47.

XX DR N-PSDB; AAI58161.

XX PT Novel nucleic acids and polypeptides, useful for treating disorders

XX PT such as central nervous system injuries -

XX PS Example 4; SEQ ID NO 2150; 10078pp; English.

XX CC The invention relates to human nucleic acids (AAI57798-AAI61369) and

XX CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,

XX CC immunosuppressant and cytostatic activity. The polynucleotides are useful

XX CC in gene therapy. A composition containing a polypeptide or polynucleotide

XX CC of the invention may be used to treat diseases of the peripheral nervous

XX CC system, such as peripheral nervous injuries, peripheral neuropathy and

XX CC localised neuropathies and central nervous system diseases, such as

XX CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

XX CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the

XX CC utilisation of the activities such as: immune system suppression,

XX CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic

XX CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,

XX CC assays for receptor activity, arthritis and inflammation, leukaemias and

XX CC C.N.S disorders.

XX CC Note: The sequence data for this patent did not form part of the printed

XX CC specification.

XX SQ Sequence 260 AA;
Query Match 100.0%; Score 1319; DB 22; Length 260;
Best Local Similarity 100.0%; Pred. No. 2.9e-104;
Matches 260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGSPLWGPAGGVGLLVLLGLFRPPPPALCARPVKEPRGLSAAAPPLAETGAPRRFR 60

Db 1 MAGSPLWGPAGGVGLLVLLGLFRPPPPALCARPVKEPRGLSAAAPPLAETGAPRRFR 60

QY 61 RSVPRGEAAGAVQELARALAHLLAEERQERARAEAOEDQOARVLAQLLRVWGAPRNSD 120

Db 61 RSVPRGEAAGAVQELARALAHLLAEERQERARAEAOEDQOARVLAQLLRVWGAPRNSD 120

QY 121 PALGLDDDDPDAPAAQALARALLRARLDPAALAAQLVPAPVPAALAPRPPVYDDGPGAGFDA 180

Db 121 PALGLDDDDPDAPAAQALARALLRARLDPAALAAQLVPAPVPAALAPRPPVYDDGPGAGFDA 180

QY 181 EEAGDETPDVPPELLRYLLGRILAGSADSEGVAAAPRRLRAADHDVGSSELPPEGVLGALL 240

Db 181 EEAGDETPDVPPELLRYLLGRILAGSADSEGVAAAPRRLRAADHDVGSSELPPEGVLGALL 240

QY 241 RVKLETPAPQVPARRLLPP 260

Db 241 RVKLETPAPQVPARRLLPP 260

RESULT 4

AAB88596

ID AAB88596 standard; Protein; 260 AA.

XX AC AAB88596;

XX DT 04-JUN-2001 (first entry)

XX DE Human hydrophobic domain containing protein clone HP10756 #100.

XX KW Human; hydrophobic domain; immunosuppressant; anti-HIV; neuroprotective;

XX KW antianaemic; vulnery; antiulcer; osteopathic; anti-inflammatory;

XX KW cytostatic; gene therapy; autoimmune disorder; multiple sclerosis;

XX KW HIV infection; anaemia; burn; ulcer; osteoporosis; tumour; wound healing;

XX KW inflammatory bowel disease; nutritional supplement; appetite; vaccine;

XX KW behavioural characteristic; immune response.

XX OS Homo sapiens.

XX PN WO200112660-A2.

XX PD 22-FEB-2001.

XX PF 10-AUG-2000; 2000WO-JP05356.

XX PR 17-AUG-1999; 99JP-0230344.

XX PR 07-SEP-1999; 99JP-0252551.

XX PR 01-OCT-1999; 99JP-0281132.

XX PR 22-OCT-1999; 99JP-0301624.

XX PR 04-NOV-1999; 99JP-0313877.

XX PA (SAGA) SAGAMI CHEM RES CENT.

XX PA (PROT-) PROTEGENE INC.

XX PI Kato S, Kimura T;

XX DR WPI; 2001-160059/16.

XX DR N-PSDB; AAF94486.

XX PT Human proteins with hydrophobic domains and the DNAs which encode them

XX PT are useful for treating autoimmune disorders, burns and tumors and for

XX PT screening novel pharmaceuticals -

XX PS Claim 1; Page 422-423; 518pp; English.

XX PS

CC AAF94417 to AAF94516 encode the human proteins given in AAB88557 to
 CC AAB88606 (I) which have a hydrophobic domain. (I) have immunosuppressant,
 CC anti-HIV, neuroprotective, antianemic, vulnerary, antitumor,
 CC osteopathic, anti-inflammatory and cytostatic activities, and can be
 CC used in gene therapy. (I) can be used as pharmaceuticals and as antigens
 CC to prepare antibodies. DNA and cDNA (II) encoding (I) can be used as
 CC probes for genetic diagnosis and gene sources for gene therapy or for
 CC producing (I) in large quantities. Cells containing (II) are used for
 CC the detection of ligands or receptors corresponding to membrane or
 CC secretory proteins and to screen small molecule novel pharmaceuticals.
 CC Antibodies directed to (I) can be used for the detection, quantification
 CC and purification of (I). Activities of (I) may include cytokine and cell
 CC proliferation/differentiation function, immune stimulating or suppressing
 CC activity, haematopoiesis regulating activity, tissue growth activity,
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, receptor/ligand activity and anti-inflammatory
 CC activity. (I) and (II) can be used to treat autoimmune disorders e.g.
 CC multiple sclerosis, HIV infections, anaemia, burns, ulcers, osteoporosis,
 CC inflammatory bowel disease and tumours. (I) and (II) can also be used for
 CC wound healing, as nutritional sources or supplements e.g. as amino acid,
 CC carbon or nitrogen source, to effect metabolism, catabolism, anabolism,
 CC processing and utilisation of dietary fat, protein, carbohydrate,
 CC vitamins and minerals, to effect behavioural characteristics, to affect
 CC appetite, and can act as antigens in vaccines to raise an immune response
 CC to the protein or another material cross-reactive with the protein.
 XX

SQ Sequence 260 AA;

Query Match 100.0%; Score 1319; DB 22; Length 260;
 Best Local Similarity 100.0%; Pred. No. 2.9e-104;
 Matches 260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGSPLLWGPRAGGVLVLLGLFRPPPCALCARPVKEPRGLSAASPPLAETGAPRRFR 60
 DB 1 MAGSPLLWGPRAGGVLVLLGLFRPPPCALCARPVKEPRGLSAASPPLAETGAPRRFR 60

QY 61 RSVPRGEAAGVQELARALAHLEAEERARAEAEQDQARVLAQLLRVWGAPRNSD 120
 DB 61 RSVPRGEAAGVQELARALAHLEAEERARAEAEQDQARVLAQLLRVWGAPRNSD 120

QY 121 PALGLDDDDPAPAAQALARALLRDLRDLPAALAAQLVPAPVPAALRPPVYDDGAPGPA 180
 DB 121 PALGLDDDDPAPAAQALARALLRDLRDLRDLPAALAAQLVPAPVPAALRPPVYDDGAPGPA 180

QY 181 EAGDETDPDPELLRYLLGRILAGSADSEGVAAAPRRLRRAADHDVGSLEPPEGVLGALL 240
 DB 181 EAGDETDPDPELLRYLLGRILAGSADSEGVAAAPRRLRRAADHDVGSLEPPEGVLGALL 240

QY 241 RVKRLTAPQVPARRLLPP 260
 DB 241 RVKRLTAPQVPARRLLPP 260

RESULT 5

AAM39006
 ID AAM39006 standard; Protein: 223 AA.

XX AAM39006;

XX 22-OCT-2001 (first entry)

XX Human polypeptide SEQ ID NO 2151.

XX Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.

XX Homo sapiens.

XX WO200153312-A1.

PN

XX 26-JUL-2001.
 PD 26-DEC-2000; 2000WO-US34263.
 PF 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX (HYSE-) HYSEQ INC.
 PA Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Weinman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX WPI; 2001-442253/47.
 DR N-PSDB; AAI58162.
 XX Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX Example 4; SEQ ID NO 2151; 10078pp; English.
 XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with neurotropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: immune system suppression,
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX Sequence 223 AA;
 SQ

Query Match 83.8%; Score 1105.5; DB 22; Length 223;
 Best Local Similarity 85.8%; Pred. No. 3.4e-86;
 Matches 223; Conservative 0; Mismatches 0; Indels 37; Gaps 1;

QY 1 MAGSPLLWGPRAGGVLVLLGLFRPPPCALCARPVKEPRGLSAASPPLAETGAPRRFR 60
 DB 1 MAGSPLLWGPRAGGVLVLLGLFRPPPCALCARPVKEPRGLSAASPPLAETGAPRRFR 60

QY 61 RSVPRGEAAGVQELARALAHLEAEERARAEAEQDQARVLAQLLRVWGAPRNSD 120
 DB 61 RSVPRGEAAGVQELARALAHLEAEERARAEAEQDQARVLAQLLRVWGAPRNSD 120

QY 121 PALGLDDDDPAPAAQALARALLRDLRDLPAALAAQLVPAPVPAALRPPVYDDGAPGPA 180
 DB 121 PALGLDDDDPAPAAQALARALLRDLRDLRDLPAALAAQLVPAPVPAALRPPVYDDGAPGPA 180

QY 181 EAGDETDPDPELLRYLLGRILAGSADSEGVAAAPRRLRRAADHDVGSLEPPEGVLGALL 240
 DB 181 EAGDETDPDPELLRYLLGRILAGSADSEGVAAAPRRLRRAADHDVGSLEPPEGVLGALL 240

QY 241 RVKRLTAPQVPARRLLPP 260
 DB 204 RVKRLTAPQVPARRLLPP 223

RESULT 6

AAY71960
 ID AAY71960 standard; Protein; 258 AA.
 XX
 AC AAY71960;
 XX
 DT 26-MAR-2001 (first entry)
 XX
 DE Mouse pituitary hormone, pituitrone.
 XX
 DE
 KW Mouse; pituitary; pituitrone; therapy; immune disorder; anaemia;
 KW Digeorge syndrome; haematopoietic cell; ataxia telangiectasia;
 KW Wiskott-Aldrich disorder; blood coagulation disorder; Addison's disease;
 KW auto-immune disorder; multiple sclerosis; systemic lupus erythematosus;
 KW SLE; hyperproliferative disorder; gene therapy; neoplasm;
 KW infectious disease; immunomodulatory; cytostatic; antimicrobial.
 XX
 OS Mus sp.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..34
 FT /label= Signal_peptide
 FT /note= "An alternative signal sequence referred in
 FT page 11"
 FT Peptide 1..33
 FT /label= Signal_peptide
 FT Protein 35..258
 FT /note= "Mature mouse pituitrone"
 FT /note= "An alternative mature protein referred in
 FT page 11"
 FT Protein 34..258
 FT /note= "Mature mouse pituitrone"
 XX
 PN WO200066778-A1.
 XX
 XX 09-NOV-2000.
 XX
 XX 27-APR-2000; 2000WO-US11211.
 XX
 XX 30-APR-1999; 99US-0131966.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Ruben SM, Ni J;
 XX
 XX WPI; 2000-687547/67.
 DR N-PSDB; AAD02076.
 XX
 XX Nucleic acids encoding a novel pituitary hormone (pituitrone), useful
 PT for diagnosing, preventing and treating e.g. immune disorders,
 PT hyperproliferative disorders and blood coagulation disorders -
 XX
 XX Claim 11; Fig 2; 277pp; English.
 PS
 CC The present sequence is a mouse pituitary hormone, pituitrone.
 CC Pituitrone is highly expressed in pituitary gland and is
 CC also expressed in brain tissues, spinal cord and kidney. It may be
 CC used as antigens in the production of antibodies against pituitrone
 CC and in assays to identify modulators. Pituitrone cDNAs are
 CC also useful in gene therapy. Pituitrone may be used in the prevention,
 CC treatment and diagnosis of diseases associated with inappropriate
 CC pituitrone expression. It may be useful in treating disorders related
 CC to reproductive and renal system, immune disorders, disorders of
 CC haematopoietic cells (e.g., anaemia, Digeorge syndrome, ataxia
 CC telangiectasia and Wiskott-Aldrich disorder), blood coagulation
 CC disorders, autoimmune disorders (e.g. Addison's disease, multiple
 CC sclerosis and systemic lupus erythematosus (SLE)), hyperproliferative
 CC disorders (e.g. neoplasms of bone, liver and pancreas) and infectious
 CC diseases.
 XX
 SQ Sequence 258 AA;

Query Match 77.3%; Score 1019; DB 21; Length 258;
 Best Local Similarity 79.2%; Pred. No. 9.2e-79;

Matches 206; Conservative 13; Mismatches 39; Indels 2; Gaps 1;
 OY 1 MAGSPLMGPRAGGVLVLLGLFRPPALCARPVKEPRGLSAAAPPLAETGAPRRFR 60
 DB 1 MAGSPLLLRAAGGGVILVLLGLLRLPPTLSARPVKEPRSLRSAPLVTSTPLRLR 60
 OY 61 RSVPRGEAAGVQELARALAHLLLEAEQERARAEAEQEQARVLAQLLRVWCAPRNSD 120
 DB 61 RAVPRGEAAGVQELARAVATVLEAEQERARAEAEQEQARVLAQLLRVWCAPRNSD 120
 OY 121 PALGLDDDDPAPAAQALRALRLRDLPAALAAQLVPAPVPAALRPPRPVYDDGPGAPDA 180
 DB 121 PPLAPDDDDPAPAAQALRALRLRDLPAALAAQLVPA--PAAAPRPVYDDGPGTGDV 178
 OY 181 EEAGDETPDVPDPELLRYLLGRILAGSADSEGVAAAPRRLRAADHDVSELPPGVLGALL 240
 DB 179 EDAGDETPDVPDPELLRYLLGRILTSSEPEAAAPRRLRSVDQDLGPEVPPENVLGALI 238
 OY 241 RVKRLTPAPQVPARRLLPP 260
 DB 239 RVKRLNPSQAPARRLLPP 258
 RESULT 7
 AAY71962
 ID AAY71962 standard; Protein; 262 AA.
 XX
 AC AAY71962;
 XX
 DT 26-MAR-2001 (first entry)
 XX
 DE Consensus sequence of human, mouse and rat pituitrone.
 XX
 KW Rat; human; mouse; pituitary; pituitrone; therapy; immune disorder;
 KW Digeorge syndrome; haematopoietic cell; ataxia telangiectasia;
 KW Wiskott-Aldrich disorder; blood coagulation disorder; Addison's disease;
 KW auto-immune disorder; multiple sclerosis; systemic lupus erythematosus;
 KW SLE; hyperproliferative disorder; gene therapy; neoplasm; anaemia;
 KW infectious disease; immunomodulatory; cytostatic; antimicrobial.
 XX
 OS Rattus sp.
 OS Mus sp.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 8..12 /label= Unknown
 FT Misc-difference 17 /label= Unknown
 FT Misc-difference 26 /label= Unknown
 FT Misc-difference 28 /label= Unknown
 FT Misc-difference 31 /label= Unknown
 FT Misc-difference 33 /label= Unknown
 FT Misc-difference 42 /label= Unknown
 FT Misc-difference 46 /label= Unknown
 FT Misc-difference 48 /label= Unknown
 FT Misc-difference 51 /label= Unknown
 FT Misc-difference 54 /label= Unknown
 FT Misc-difference 55 /label= Unknown
 FT Misc-difference 57 /label= Unknown
 FT Misc-difference 59 /label= Unknown
 FT /label= Unknown

FT Protein 34, 148
FT /label= Mature_human_secreted_protein
PN WO200151504-A1.
XX 19-JUL-2001.
XX 12-JAN-2001; 2001WO-US00911.
XX 13-JAN-2000; 2000US-0482273.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Ruben SM, Komatsoulis GA, Duan DR, Rosen CA, Moore PA, Shi Y;
PI Lafleur DW, Olsen H, Brewer LA, Florence KA, Young PE, Soppet DR;
PI Endress GA, Mucenski M, Ebner R;
XX WPI; 2001-425865/45.
DR N-PSDB; AAD11677.
XX Isolated nucleic acid molecule encoding a human secreted protein is
PT used in preventing, treating or ameliorating a medical condition -
PT Claim 11; Page 773-774; 864pp; English.
PS AAD11630-AAD11721 represent cDNAs corresponding to 71 human secreted
CC protein genes, and AAE06041-AAE06132 represent the proteins they encode.
CC AAE06133-AAE06205 represent human secreted protein fragments.
CC The secreted proteins and their genes are useful for preventing, treating
CC or ameliorating medical conditions, e.g., by protein or gene therapy.
CC Pathological conditions can be diagnosed by determining the amount of the
CC new protein in a sample or by determining the presence of mutations in
CC the new genes. Specific uses are described for each of the 71 genes,
CC based on the tissues in which they are most highly expressed, and include
CC developing products for the diagnosis or treatment of proliferative
CC disorders, cancer, tumours, foetal and developmental abnormalities,
CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune
CC diseases (e.g., rheumatoid arthritis), inflammation, allergies, and
CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),
CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,
CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,
CC angiotensin-related disorders, kidney disorders, gastrointestinal disorders,
CC pregnancy-related disorders, endocrine disorders, and infectious. The
CC proteins can also be used to aid wound healing and epithelial cell
CC proliferation, to prevent skin aging due to sunburn, to maintain organs
CC before transplantation, for supporting cell culture of primary tissues,
CC to regenerate tissues, to identify their cognate ligands or binding
CC partners, and in chemotaxis, and can be used as a food additive or
CC preservative to modify storage properties. Antibodies specific for a
CC protein of the invention can be used in alleviating symptoms associated
CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,
CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA).
CC The present sequence represents a human secreted protein of
CC the invention.
XX Sequence 148 AA;
SQ Query Match 23.4%; Score 308; DB 22; Length 148;
Best Local Similarity 50.3%; Pred. No. 1.4e-18;
Matches 88; Conservative 3; Mismatches 34; Indels 50; Gaps 8;
QY 1 MAGSPLWGPAGVGLVLLGLRPPPPALCARPVKEPRGISAASPPLAETG----- 54
DB 1 MAGSPLWGPAGVGLVLLGLRPPPPALCARPVKEPRGISAASPPPLARLALLAASG 60
QY 55 --APRRFRSVPV-GBAAGVQELARALAHLEAQRERARAEQAEQDQARVLAQLR 111
DB 61 GQCFVRRRCRCRPGAGAGA-----SAGAEQERARAEQAR-----LR 98
QY 112 V-----WGAPRNSDPALGDDDDPAPAAQLAR-----ALLRRLDPALAAQLWPA 157
DB 99 ISRRASW-----RSCCAGS-----APPATLRLWAWTTTPTLRQSSSLACSAPA 143

RESULT 9
ABG33910
ID ABG33910 standard; Protein; 148 AA.
XX ABG33910;
XX 15-JUL-2002 (first entry)
XX Human secreted protein encode by gene 48.
XX Human; secreted protein; gene therapy; immunosuppressive;
KW antarthritic; antirheumatic; antiproliferative; cytostatic; cardiantic;
KW vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
KW virucide; fungicide; ophthalmological; autoimmune disease; neoplasm;
KW rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;
KW cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;
KW angiogenesis; nervous system disorder; Alzheimer's disease; infection;
KW ocular disorder; corneal infection; wound healing; skin aging;
KW epithelial cell proliferation; food additive.
XX Homo sapiens.
XX WO200226931-A2.
XX 04-APR-2002.
XX 24-SEP-2001; 2001WO-US29871.
XX 25-SEP-2000; 2000US-234925P.
PR 12-JAN-2001; 2001WO-US00911.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Ruben SM, Komatsoulis G, Duan DR, Rosen CA, Moore PA, Shi Y;
PI Lafleur DW, Olsen H, Brewer LA, Florence KA, Young PE, Soppet DR;
PI Endress GA, Mucenski M, Ebner R;
XX WPI; 2002-362489/39.
DR N-PSDB; ABK69773.
XX Novel 71 isolated secreted polypeptides and polynucleotides encoding
PT the polypeptides, useful for treating Huntington's disease, sepsis,
PT meningitis, thrombocytopaenia, haemolytic anaemia, rheumatoid arthritis,
PT asthma -
XX Claim 11; Page 1252-1253; 1478pp; English.
XX The invention relates to an isolated nucleic acid molecule (or its
CC fragment, homologue complement or allelic variant) encoding a human
CC secreted protein (and its fragment, domain, epitope, variant, secreted
CC form and species variant). Also included are a recombinant vector
CC comprising the nucleic acid, a recombinant host cell comprising the
CC vector, an antibody against the secreted protein, a recombinant host cell
CC that expresses the secreted protein and a method of identifying a binding
CC partner of the secreted protein. The nucleic acid and protein are used to
CC prevent, diagnose, treat or ameliorate a medical condition in e.g.
CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep
CC for example autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi and
CC ocular disorders e.g. corneal infection. Many other diseases and
CC disorders are listed in the specification. The polypeptides can also be
CC used to aid wound healing an epithelial cell proliferation, to prevent
CC skin aging due to sunburn, to maintain organs before transplantation, for
CC supporting cell culture of primary tissues, to regenerate tissues and in
CC chemotaxis. The polypeptides can also be used as a food additive or
CC preservative to increase or decrease storage capabilities. The present
CC sequence represents a novel human secreted protein of the invention.
XX Sequence 148 AA;
SQ

Query Match 23.4%; Score 308; DB 23; Length 148;
 Best Local Similarity 50.3%; Pred. No. 1.4e-18;
 Matches 88; Conservative 3; Mismatches 34; Indels 50; Gaps 8;

QY 1 MAGSPLLMGPRAGGVLVLLGLLGLFRPPPPALCARPVKEPRGLSAASPPPLAETG----- 54
 DB 1 MAGSPLLMGPRAGGVLVLLGLLGLFRPPPPALCARPVKEPRGLSAASPPPLAETG 60

QY 55 --APRRFRSVPK-GEAAGAVQELARALAHLEAEQRARAEAEQEDQOARVLAQLLR 111
 DB 61 GQCPVRRGRRCRCPAGAGA-----SAGAERQERARAEQR-----LR 98

QY 112 V-----WGAPRNSDPALGLDDDDPAPAAQLAR-----ALLRLRDLPAALAAQLVPA 157
 DB 99 ISRRASW-----RSCCAGS-----APPATLRLWAWTTTPTLRQSSSLALCSAPA 143

RESULT 10
 AAY87111 standard; Protein: 149 AA.
 XX AC AAY87111;
 DT 09-MAY-2000 (first entry)
 DE Human secreted protein sequence SEQ ID NO:150.
 KW Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
 KW antiinflammatory; nontropic; neuroprotective; antiallergic; cancer;
 KW tumour; neurodegenerative disorder; developmental abnormality; allergy;
 KW foetal deficiency; blood disorder; immune system disorder; arthritis;
 KW autoimmune disease; hepatic disease; renal disease; inflammation;
 KW Alzheimer's disease; behavioural disorder; schizophrenia; osteoporosis;
 KW infection; AIDS; spinal cord injury; transplant rejection; diabetes;
 KW asthma; sepsis; acne; psoriasis; cardiovascular disorder;
 KW reproductive disorder; gastrointestinal disorder; respiratory disorder;
 KW metabolic disorder; food additive; preservative.
 XX OS Homo sapiens.
 XX PN WO200004140-A1.
 XX PD 27-JAN-2000.
 XX PF 14-JUL-1999; 99WO-US15849.
 XX PR 15-JUL-1998; 98US-0092921.
 XX PR 15-JUL-1998; 98US-0092922.
 XX PR 15-JUL-1998; 98US-0092956.
 XX PA (HUMA-) HUMAN GENOME SCI INC.
 XX PI Ruben SM, Komatsoulis G, Duan RD, Rosen CA, Moore PA, Shi Y;
 PI Lafleur DW, Ebner R, Olsen HS, Brewer LA, Florence KA, Young PE;
 XX Mucenski M, Dresser GA, Soppet DR;
 XX WPI; 2000-161128/14.
 XX N-PSDB; AAZ98064.
 XX PT New isolated human genes, useful for diagnosis and treatment of, e.g.
 XX PT cancers, neurological or blood disorders -
 XX PS Claim 11; Page 406-407; 494pp: English.
 XX CC The polynucleotide sequences given in AAZ98017 to AAZ98108 encode the
 XX CC human secreted proteins given in AAY87064 to AAY87223. Human secreted
 XX CC protein can have activities based on the tissues and cells the genes are
 XX CC expressed in. Examples of activities include: cytostatic;
 XX CC immunosuppressive; antiinflammatory; nontropic; neuroprotective; and
 XX CC antiallergic. The polynucleotides and their corresponding secreted
 XX CC polypeptides are useful for preventing, treating or ameliorating medical
 XX CC conditions, e.g. by protein or gene therapy. Also pathological conditions

CC can be diagnosed by determining the amount of the new polypeptides in a
 CC sample or by determining the presence of mutations in the new
 CC polynucleotides. Human secreted protein s and their polynucleotides can
 CC be used for developing products for the diagnosis or treatment of cancer,
 CC tumours, neurodegenerative disorders, developmental abnormalities and
 CC foetal deficiencies, blood disorders, diseases of the immune system,
 CC autoimmune diseases, hepatic and renal disease, inflammation,
 CC allergies, Alzheimer's disease, behavioural disorders, schizophrenia,
 CC osteoporosis, arthritis, infections, asthma, sepsis, acne, psoriasis,
 CC transplant rejection, diabetes, AIDS, spinal cord injuries,
 CC cardiovascular disorders, reproductive disorders, gastrointestinal
 CC disorders, respiratory disorders and metabolic disorders. The
 CC proteins or polynucleotides can also be used as food additives or
 CC preservatives. The proteins are also useful for identifying their
 CC binding partners. AAZ98008 to AAZ98016 and AAY87063 are sequence used in
 CC the exemplification of the present invention.
 XX SQ Sequence 149 AA;

Query Match 23.4%; Score 308; DB 21; Length 149;
 Best Local Similarity 50.3%; Pred. No. 1.4e-18;
 Matches 88; Conservative 3; Mismatches 34; Indels 50; Gaps 8;

QY 1 MAGSPLLMGPRAGGVLVLLGLLGLFRPPPPALCARPVKEPRGLSAASPPPLAETG----- 54
 DB 1 MAGSPLLMGPRAGGVLVLLGLLGLFRPPPPALCARPVKEPRGLSAASPPPLAETG 60

QY 55 --APRRFRSVPK-GEAAGAVQELARALAHLEAEQRARAEAEQEDQOARVLAQLLR 111
 DB 61 GQCPVRRGRRCRCPAGAGA-----SAGAERQERARAEQR-----LR 98

QY 112 V-----WGAPRNSDPALGLDDDDPAPAAQLAR-----ALLRLRDLPAALAAQLVPA 157
 DB 99 ISRRASW-----RSCCAGS-----APPATLRLWAWTTTPTLRQSSSLALCSAPA 143

RESULT 11
 ID AAY87199 standard; Protein: 166 AA.
 XX AC AAY87199;
 XX DT 09-MAY-2000 (first entry)
 XX DE Human secreted protein sequence SEQ ID NO:238.
 KW Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
 KW antiinflammatory; nontropic; neuroprotective; antiallergic; cancer;
 KW tumour; neurodegenerative disorder; developmental abnormality; allergy;
 KW foetal deficiency; blood disorder; immune system disorder; arthritis;
 KW autoimmune disease; hepatic disease; renal disease; inflammation;
 KW Alzheimer's disease; behavioural disorder; schizophrenia; osteoporosis;
 KW infection; AIDS; spinal cord injury; transplant rejection; diabetes;
 KW asthma; sepsis; acne; psoriasis; cardiovascular disorder;
 KW reproductive disorder; gastrointestinal disorder; respiratory disorder;
 KW metabolic disorder; food additive; preservative.
 XX OS Homo sapiens.
 XX PN WO200004140-A1.
 XX PD 27-JAN-2000.
 XX PF 14-JUL-1999; 99WO-US15849.
 XX PR 15-JUL-1998; 98US-0092921.
 XX PR 15-JUL-1998; 98US-0092922.
 XX PR 15-JUL-1998; 98US-0092956.
 XX PA (HUMA-) HUMAN GENOME SCI INC.
 XX PI Ruben SM, Komatsoulis G, Duan RD, Rosen CA, Moore PA, Shi Y;
 PI Lafleur DW, Ebner R, Olsen HS, Brewer LA, Florence KA, Young PE;

PI Mucenski M, Endress GA, Soppet DR;
XX WPI; 2000-161128/14.
DR N-PSDB; AA298064.
XX
PT New isolated human genes, useful for diagnosis and treatment of, e.g.
PT cancers, neurological or blood disorders
XX
XX Disclosure; Page 470; 494pp; English.
XX
CC The polynucleotide sequences given in AA298017 to AA298108 encode the
CC human secreted proteins given in AA298064 to AA298723. Human secreted
CC protein can have activities based on the tissues and cells the genes are
CC expressed in. Examples of activities include: cytostatic;
CC immunosuppressive; antiinflammatory; neurotropic; neuroprotective; and
CC anti-allergic. The polynucleotides and their corresponding secreted
CC polypeptides are useful for preventing, treating or ameliorating medical
CC conditions, e.g. by protein or gene therapy. Also pathological conditions
CC can be diagnosed by determining the amount of the new polypeptides in a
CC sample or by determining the presence of mutations in the new
CC polynucleotides. Human secreted proteins and their polynucleotides can
CC be used for developing products for the diagnosis or treatment of cancer,
CC tumours, neurodegenerative disorders, developmental abnormalities and
CC foetal deficiencies, blood disorders, diseases of the immune system,
CC autoimmune diseases, hepatic and renal disease, inflammation,
CC allergies, Alzheimer's disease, behavioural disorders, schizophrenia,
CC osteoporosis, arthritis, infections, AIDS, spinal cord injuries,
CC transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,
CC cardiovascular disorders, reproductive disorders, gastrointestinal
CC disorders, respiratory disorders and metabolic disorders. The
CC proteins or polynucleotides can also be used as food additives or
CC preservatives. The proteins are also useful for identifying their
CC binding partners. AA298008 to AA298016 and AA298063 are sequence used in
CC the exemplification of the present invention.
XX
XX Sequence 166 AA;
XX
Query Match 23.4%; Score 308; DB 21; Length 166;
Best Local Similarity 50.3%; Pred. No. 1.6e-18;
Matches 88; Conservative 3; Mismatches 34; Indels 50; Gaps 8;
QY 1 MAGSPLWGPAGVGLVLLVLLGLFRPPALCARPVKEPGLSAASPPLAETG----- 54
Db 19 MAGSPLWGPAGVGLVLLVLLGLFRPPALCARPVKEPGLSAASPPLARLALLAASG 78
QY 55 --APRRFRSVP-GEAAGAVQELARALAHLEAERQERARAQAEDQOARVLAQLR 111
Db 79 GQCPVRRRCRCPGAGAG-----SAGAERQERARAQA-----LR 116
QY 112 V-----WGAPRNSDPALGLDDDDPAPAAQLAR-----ALLRARLDPAALAAQLVPA 157
Db 117 ISRRASW-----RSCCAGS-----APPATLRLWAWTTTTRLQRSSSLALCSAPA 161
RESULT 12
AAE06176
XX AAE06176 standard; Protein; 166 AA.
XX
XX AAE06176;
XX
XX 24-SEP-2001 (first entry)
XX
XX Human gene 48 encoded secreted protein fragment, SEQ ID NO:238.
XX
XX Human; secreted protein; proliferative disorder; cancer; tumour; asthma;
XX foetal abnormality; developmental abnormality; haematopoietic disorder;
XX immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
XX Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;
XX psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;
XX inflammation; neurological disorder; Alzheimer's disease; food additive;
XX angiogenic disorder; kidney disorder; gastrointestinal disorder; allergy;
XX pregnancy-related disorder; endocrine disorder; infection; wound healing;
XX cell culture; chemotaxis; vulnery; binding partner identification;

XX gene therapy.
XX Homo sapiens.
XX WO200151504-A1.
XX 19-JUL-2001.
XX
XX 12-JAN-2001; 2001WO-US00911.
XX
XX 13-JAN-2000; 2000US-0482273.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Komatsoulis GA, Duan DR, Rosen CA, Moore PA, Shi Y;
XX Lafleur DW, Olsen HS, Brewer LA, Florence KA, Young PE, Soppet DR;
XX Endress GA, Mucenski M, Ebner R;
XX WPI; 2001-425865/45.
XX
XX Isolated nucleic acid molecule encoding a human secreted protein is
XX used in preventing, treating or ameliorating a medical condition -
XX
XX Disclosure; Page 122; 864pp; English.
XX
XX AAD11630-AAD11721 represent cDNAs corresponding to 71 human secreted
XX protein genes, and AAE06041-AAE06132 represent the proteins they encode.
XX AAE06133-AAE06205 represent human secreted protein fragments.
XX The secreted proteins and their genes are useful for preventing, treating
XX or ameliorating medical conditions, e.g., by protein or gene therapy.
XX Pathological conditions can be diagnosed by determining the amount of the
XX new protein in a sample or by determining the presence of mutations in
XX the new genes. Specific uses are described for each of the 71 genes,
XX based on the tissues in which they are most highly expressed, and include
XX developing products for the diagnosis or treatment of proliferative
XX disorders, cancer, tumours, foetal and developmental abnormalities,
XX haematopoietic disorders, diseases of the immune system, AIDS, autoimmune
XX diseases (e.g., rheumatoid arthritis), inflammation, allergies,
XX neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),
XX cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,
XX psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,
XX angiogenic disorders, kidney disorders, gastrointestinal disorders,
XX pregnancy-related disorders, endocrine disorders, and infections. The
XX proteins can also be used to aid wound healing and epithelial cell
XX proliferation, to prevent skin aging due to sunburn, to maintain organs
XX before transplantation, for supporting cell culture of primary tissues,
XX to regenerate tissues, to identify their cognate ligands or binding
XX partners, and in chemotaxis, and can be used as a food additive or
XX preservative to modify storage properties. Antibodies specific for a
XX protein of the invention can be used in alleviating symptoms associated
XX with the disorders mentioned above, and in diagnostic immunoassays e.g.,
XX radioimmunoassay or enzyme linked immunosorbent assay (ELISA).
XX The present sequence represents a human secreted protein fragment
XX referred to in the disclosure of the invention.
XX
XX Sequence 166 AA;
XX
Query Match 23.4%; Score 308; DB 22; Length 166;
Best Local Similarity 50.3%; Pred. No. 1.6e-18;
Matches 88; Conservative 3; Mismatches 34; Indels 50; Gaps 8;
QY 1 MAGSPLWGPAGVGLVLLVLLGLFRPPALCARPVKEPGLSAASPPLAETG----- 54
Db 19 MAGSPLWGPAGVGLVLLVLLGLFRPPALCARPVKEPGLSAASPPLARLALLAASG 78
QY 55 --APRRFRSVP-GEAAGAVQELARALAHLEAERQERARAQAEDQOARVLAQLR 111
Db 79 GQCPVRRRCRCPGAGAG-----SAGAERQERARAQA-----LR 116
QY 112 V-----WGAPRNSDPALGLDDDDPAPAAQLAR-----ALLRARLDPAALAAQLVPA 157
Db 117 ISRRASW-----RSCCAGS-----APPATLRLWAWTTTTRLQRSSSLALCSAPA 161

RESULT 13
ABG34000
ID ABG34000 standard; Protein; 166 AA.
XX
XX
AC ABG34000;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human secreted protein #2 encoded by gene 48.
XX
KW Human; secreted protein; gene therapy; immunosuppressive;
KW antichronic; antineumatic; antiproliferative; cytostatic; cardiant;
KW vasotropic; cerebroprotective; nootropic; neuroproteic; antibacterial;
KW virucide; fungicide; ophthalmological; autoimmune disease; neoplasm;
KW rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;
KW cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;
KW angiogenesis; nervous system disorder; Alzheimer's disease; infection;
KW ocular disorder; corneal infection; wound healing; skin aging;
KW epithelial cell proliferation; food additive.
XX
OS Homo sapiens.
XX
PN WO200226931-A2.
XX
PD 04-APR-2002.
XX
XX 24-SEP-2001; 2001WO-US29871.
XX
XX 25-SEP-2000; 2000US-234925P.
PR 12-JAN-2001; 2001WO-US00911.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Komatsoulis G, Duan DR, Rosen CA, Moore PA, Shi Y;
PI Lafleur DW, Olsen H, Brewer LA, Florence KA, Young PE, Soppet DR;
PI Endress GA, Mucenski M, Ebner R;
XX
XX WPI; 2002-362489/39.
DR N-PSDB; ABK69773.
XX
XX Novel 71 isolated secreted polypeptides and polynucleotides encoding
PT the polypeptides, useful for treating Huntington's disease, sepsis,
PT meningitis, thrombocytopaenia, haemolytic anaemia, rheumatoid arthritis,
PT asthma
XX
XX Disclosure; Page 1316-1317; 1478pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (or its
CC fragment, homologue complement or allelic variant) encoding a human
CC secreted protein (and its fragment, domain, epitope, variant, secreted
CC form and species variant). Also included are a recombinant vector
CC comprising the nucleic acid, a recombinant host cell comprising the
CC vector, an antibody against the secreted protein, a recombinant host cell
CC that expresses the secreted protein and a method of identifying a binding
CC partner of the secreted protein. The nucleic acid and protein are used to
CC prevent, diagnose, treat or ameliorate a medical condition in e.g.
CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep
CC for example autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi and
CC ocular disorders e.g. corneal infection. Many other diseases and
CC disorders are listed in the specification. The polypeptides can also be
CC used to aid wound healing an epithelial cell proliferation, to prevent
CC skin aging due to sunburn, to maintain organs before transplantation, for
CC supporting cell culture of primary tissues, to regenerate tissues and in
CC chemotaxis. The polypeptides can also be used as a food additive or
CC preservative to increase or decrease storage capabilities. The present
CC sequence represents a novel human secreted protein of the invention.
XX
SQ Sequence 166 AA;

Query Match 23.4%; Score 308; DB 23; Length 166;
Best Local Similarity 50.3%; Pred. No. 1.6e-18;
Matches 88; Conservative 3; Mismatches 34; Indels 50; Gaps 8;
QY 1 MAGSPLWGPAGGVLVLLVLLGLFRPPPPALCARPVKEPGLSRAASPPLAETG----- 54
DB 19 MAGSPLWGPAGGVLVLLVLLGLFRPPPPALCARPVKEPGLSRAASPPPLALILAASG 78
QY 55 --APRRERSVPR-GEAGAGVQELARALAHLEAEQERARAEAEQARVLAQLLR 111
DB 79 GOCPEVRRRCRCPGAGAGA-----SAGAEQERARAEQOR-----LR 116
QY 112 V-----WGAPRNSDPALGLDDPDPAQAQLAR-----ALLRLRDLPAALAAQLVPA 157
DB 117 ISRRASW-----RSCCAGS-----APPATLRLWAWTTTPTLRQSSSLALCSAPA 161
RESULT 14
AAV71961
ID AAV71961 standard; Protein; 69 AA.
XX
XX AC AAV71961;
XX
DT 26-MAR-2001 (first entry)
XX
DE Rat pituitary hormone, pituitrone.
XX
KW Rat; pituitary; pituitrone; therapy; immune disorder; anaemia;
KW Digeeorge syndrome; haematopoietic cell; ataxia telangiectasia;
KW Wiskott-Aldrich disorder; blood coagulation disorder; Addison's disease;
KW autoimmune disorder; multiple sclerosis; systemic lupus erythematosus;
KW SLE; hyperproliferative disorder; gene therapy; neoplasm;
KW infectious disease; immunomodulatory; cytostatic; antimicrobial.
XX
OS Rattus sp.
XX
XX WO200066778-A1.
PN
PD 09-NOV-2000.
XX
XX 27-APR-2000; 2000WO-US11211.
PF
XX 30-APR-1999; 99US-0131966.
PR
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX Ruben SM, Ni J;
PI
XX WPI; 2000-687547/67.
DR N-PSDB; AAD02077.
XX
XX Nucleic acids encoding a novel pituitary hormone (pituitrone), useful
PT for diagnosing, preventing and treating e.g. immune disorders,
PT hyperproliferative disorders and blood coagulation disorders -
XX
XX Claim 11; Fig 2; 277pp; English.
XX
XX The present sequence is a rat pituitary hormone, pituitrone.
CC Pituitrone is highly expressed in pituitary gland and is
CC also expressed in brain tissues, spinal cord and kidney. It may be
CC used as antigens in the production of antibodies against pituitrone
CC and in assays to identify modulators. Pituitrone cDNAs are
CC also useful in gene therapy. Pituitrone may be used in the prevention,
CC treatment and diagnosis of diseases associated with inappropriate
CC pituitrone expression. It may be useful in treating disorders related
CC to reproductive and renal system, immune disorders, disorders of
CC haematopoietic cells (e.g., anaemia, Digeeorge syndrome, ataxia
CC telangiectasia and Wiskott-Aldrich disorder), blood coagulation
CC disorders, autoimmune disorders (e.g. Addison's disease, multiple
CC sclerosis and systemic lupus erythematosus (SLE)), hyperproliferative
CC disorders (e.g. neoplasms of bone, liver and pancreas) and infectious
CC diseases.

CC	vaccine, cytotoxic T-cells and antibodies are induced which recognise and destroy or eliminate cells expressing IA proteins. The genes are also for generating animal models of lymphoma. This sequence represents a lymphoma associated protein described in the invention.
xx	
xx	
SQ	Sequence 388 AA;
	Query Match 10.3%; Score 136.5; DB 23; Length 388; Best Local Similarity 27.5%; pred. No. 0.0016; Matches 67; Conservative 19; Mismatches 89; Indels 69; Gaps 10;
QY	47 SPPIAETGAPRRFRSRVPRGEAGAVOELARALAHLEAEQRERARAQAEGAEQQQARVL 106 : : : : : :
Db	26 SPPIALDGPPTK-----VSGAPD-----KRERAERPVEEAAEMEGAADA 66 : : : : : :
QY	107 AQLLRV---WGAP-----RNSDPALG-LDDDPDPAQAOLARALLRLRL 145 : : : : : :
Db	67 AEGKVSPGYGSVAAGAASADTAAARAAPAADPDSGATPEDPDSTGAP-----ADP 119 : : : : : :
QY	146 DPAALAQLVPAPYPAAALRALRPVPDYDGPAAGDAEEAGDETPVDPELLRYLLGRILAG 205
Db	120 DSGFAADPDSCGAAPAAPADPDSCAAPADPADPDSCGAAPADAPDPD-----AG 167
QY	206 SADSEGVAAPRRLLRAADH-----DVGSELPPEGVIGLALLRVKRLKETPAPOVPARR --- 256 :
Db	168 AA-PEAPAAPAAETAERAHVAAPADGAPTAPAAASATRAAQVRRAASAAPASGARRKIH 226 :
QY	257 LLPP 260
Db	227 LRPP 230

Search completed: April 4, 2003, 14:16:20
Job time : 39 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 12, 2003, 18:34:13 ; Search time 274 Seconds
(without alignments)
8046.374 Million cell updates/sec

Title: US-09-803-589-5

Perfect score: 979

Sequence: 1 gaattcggcagcaggccagc.....gaggatccctaccctctggc 979

Scoring table:

Gapop 60.0 , Capext 60.0

Searched: 2185239 seqs, 1125999159 residues

Word size : 15

Total number of hits satisfying chosen parameters: 8227

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	979	100.0	979	20	AAV68668 Nucleotide sequenc
2	891	91.0	1030	22	AAI58161 Human polynucleoti
3	891	91.0	1041	22	AAI59947 Human polynucleoti
4	891	91.0	1041	22	AAI59948 Human polynucleoti
5	891	91.0	1043	21	AAAD02075 cDNA encoding huma
6	840	85.8	998	22	AAF94496 Human hydrophobic
7	678	69.3	780	22	AAF94486 Human hydrophobic
8	583	59.6	1050	21	AAZ98064 Human secreted pro
9	583	59.6	1050	22	AAAD11677 Human secreted pro

10	583	59.6	1050	24	ABK69773 Human secreted pro
11	568	58.0	919	22	AAI58162 Human polynucleoti
12	375	38.3	4503	24	ABK69876 Human secreted pro
13	226	23.1	301	24	ABN95203 Gene #1701 used to
14	226	23.1	301	24	ABL68387 Kidney cancer rela
15	60	6.1	60	24	ABN42141 Human spliced tran
16	32	3.3	991	21	ABN42141 cDNA encoding mus
17	24	2.5	652	24	ABQ31156 Oligonucleotide fo
18	24	2.5	652	24	ABQ31157 Oligonucleotide fo
19	23	2.3	652	24	ABQ31154 Oligonucleotide fo
20	23	2.3	652	24	ABQ31155 Oligonucleotide fo
21	22	2.2	425	22	AAF65901 Novel human polynu
22	22	2.2	1507	20	AAH84446 Mouse brain CNG-3
23	21	2.1	422	18	AAH88808 Nuclear steroid ho
24	21	2.1	533	22	AAAD03545 Human EST #3 to pr
25	20	2.0	355	22	AAF65178 Novel human polynu
26	20	2.0	396	21	AAAD02077 cDNA encoding rat
27	20	2.0	742	20	AAZ15143 Human gene express
28	19	1.9	156	23	AA565491 DNA encoding novel
29	19	1.9	156	23	AA582591 DNA encoding novel
30	19	1.9	156	24	AA563043 Cell death protect
31	19	1.9	251	19	AAH10968 Human biallelic po
32	19	1.9	255	20	AAH87627 Human single nucle
33	19	1.9	300	20	AAZ15019 Human gene express
34	19	1.9	349	24	AA561708 Lung small cell ca
35	19	1.9	380	24	ABK62892 Rat sequence diffe
36	19	1.9	383	22	AAF64866 Novel human polynu
37	19	1.9	389	22	AAF64961 Novel human polynu
38	19	1.9	398	22	AAF64559 Novel human polynu
39	19	1.9	425	22	AAF64749 Novel human polynu
40	19	1.9	463	23	AA565492 DNA encoding novel
41	19	1.9	464	24	AA561489 Lung small cell ca
42	19	1.9	467	20	AAV64423 Mouse developing l
43	19	1.9	476	24	ABN95942 Gene #2440 used to
44	19	1.9	486	19	AAV40148 DNA sequence (crtE
45	19	1.9	489	16	AAQ99486 3 hydroxy-beta-ion

ALIGNMENTS

RESULT 1
AAV68668
ID AAV68668 standard; DNA; 979 BP.
AC AAV68668;
DT 03-JUN-1999 (first entry)
XX Nucleotide sequence of the human Tango-81 gene.
DE Human; Tango-81; host cell; recombinant protein; antibody;
KW receptor; specific binding agent; probe; primer; hybridisation;
KW amplification; mutation; genetic mapping; ss.
XX Homo sapiens.
XX Key Location/Qualifiers
XX CDS 58..840
FT /*tag= a
FT /product= "Tango-81"
XX
XX WO9906427-A1.
XX
XX PD 11-FEB-1999.
XX
XX PF 04-AUG-1998; 98WO-US16241.
XX
XX PR 04-AUG-1997; 97US-0054645.
XX
XX PA (MILL-) MILLENNIUM BIOTHEAPEUTICS INC.
XX
XX PI McCarthy SA;

PT such as central nervous system injuries -
XX Claim 1; SEQ ID NO 364; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAI42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 1030 BP; 134 A; 404 C; 352 G; 140 T; 0 other;
Query Match 91.0%; Score 891; DB 22; Length 1030;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 941; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 38 GCCCGGCTCGCTGGGAGCAGATGCGGGGTGCGCGGTGCTTGGGGCGCGGGCGGGG 97
DB 53 GCCCGGCTCGCTGGGAGCAGATGCGGGGTGCGCGGTGCTTGGGGCGCGGGCGGGG 112
QY 98 GCGTCGGCTTTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 157
DB 113 GCGTCGGCTTTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 172
QY 158 CGCGCGCGGTAAAGAGAGCCCGCGGCTTAAAGCAGCGCTTCCGCCCTTGGCTGACACTG 217
DB 173 CGCGCGCGGTAAAGAGAGCCCGCGGCTTAAAGCAGCGCTTCCGCCCTTGGCTGACACTG 232
QY 218 GCGCTCTCTCGCCCTTCCGCGGCTCAGTCCCGAGGTGAGCGGGCGGGCGGCTGCAGG 277
DB 233 GCGCTCTCTCGCCCTTCCGCGGCTCAGTCCCGAGGTGAGCGGGCGGGCGGCTGCAGG 292
QY 278 AGCTGGCGGGCGCTGGGCGCATCTGCTGGAGCCGAAAGTCAAGAGCGGGCGGGCGGCGG 337
DB 293 AGCTGGCGGGCGCTGGGCGCATCTGCTGGAGCCGAAAGTCAAGAGCGGGCGGGCGGCGG 352
QY 338 AGCGCAGGAGGCTGAGGATCAGCAGCGCGCTCTGCGCAGCTGCTGCGCGTCTGGG 397
DB 353 AGCGCAGGAGGCTGAGGATCAGCAGCGCGCTCTGCGCAGCTGCTGCGCGTCTGGG 412
QY 398 GCGCCCGCCCGCAACTCTGATCCGGCTCTGGGCTTGGACGACGACCCCGAGCGGCTGCAG 457
DB 413 GCGCCCGCCCGCAACTCTGATCCGGCTCTGGGCTTGGACGACGACCCCGAGCGGCTGCAG 472
QY 458 CGCAGCTCGCTCGCTCTGCTCGCGCGCGCTTGAACCTTGCAGCGCTTGCAGCGCCAGC 517
DB 473 CGCAGCTCGCTCGCTCTGCTCGCGCGCGCTTGAACCTTGCAGCGCTTGCAGCGCCAGC 532
QY 518 TTGTCCCGCGCGCGCTCCCGCGCGCGCTCCGACCCCGCGCGCTTACGACGACG 577
DB 533 TTGTCCCGCGCGCGCTCCCGCGCGCGCTCCGACCCCGCGCGCTTACGACGACG 592
QY 578 GCCCGCGGGCCCGGATGCTGAGAGGAGCGGCGACGACACCCGACGCTGCAGCCCGGAGC 637
DB 593 GCCCGCGGGCCCGGATGCTGAGAGGAGCGGCGACGACACCCGACGCTGCAGCCCGGAGC 652
QY 638 TGTGAGGTACTTGTGAGGAGCGGATCTTGGCGGAGAGCGGCGACTCCGAGGGGGTGGCAG 697
DB 653 TGTGAGGTACTTGTGAGGAGCGGATCTTGGCGGAGAGCGGCGACTCCGAGGGGGTGGCAG 712
QY 698 CCGCGCGCGCTCCCGCGCGCGCGACCATGCTGGCTTGGCTTGGCTTGGCTTGGCTTGG 757

DB 713 CCGCGCGCGCTTCCCGCGCGCTCCGACACAGATGTGGGCTCTGAGTCCCGCTGAGG 772
QY 758 GCGTCTGGGGCGCTGCTGCTGTGTAAGCGCTAGAGACCCCGGGCGCCAGGTGCTG 817
DB 773 GCGTCTGGGGCGCTGCTGCTGTGTAAGCGCTAGAGACCCCGGGCGCCAGGTGCTG 832
QY 818 CAGCGCGCTTGTGCGACCTGAGCACTGCGCGGATCCCGTGCACCTGGGAGCCAGAG 877
DB 833 CAGCGCGCTTGTGCGACCTGAGCACTGCGCGGATCCCGTGCACCTGGGAGCCAGAG 892
QY 878 TGCGCCCGCGCATCCCGCACCGAGCTGCTCCCGCGCAGCAGCTCCAGAGCAACTTACCC 937
DB 893 TGCGCCCGCGCATCCCGCACCGAGCTGCTCCCGCGCAGCAGCTCCAGAGCAACTTACCC 952
QY 938 CGCGCAGCGAGCGCTTCCACCCGAGGATCCCTACCCCGCTGGC 979
DB 953 CGCGCAGCGAGCGCTTCCACCCGAGGATCCCTACCCCGCTGGC 994
RESULT 3
AAI59947/c
ID AAI59947 standard; cDNA; 1041 BP.
XX
AC AAI59947;
XX
DT 22-OCT-2001 (first entry)
XX
TX Human polynucleotide SEQ ID NO 3936.
DE
DE Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
(HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
DR P-PSDB; AAM40791.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Claim 1; SEQ ID NO 3936; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAI42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and

[illegible]

RESULT 5

AA02075
ID AAD02075 standard; cDNA; 1043 BP.
XX
AC AAD02075;
XX
DT 26-MAR-2001 (first entry)
XX
DE cDNA encoding human pituitary hormone, pituitrone.
XX
KW Human; pituitary; pituitrone; therapy; immune disorder; anaemia;
KW Digeorge syndrome; haematopoietic cell; ataxia telangiectasia;
KW Wiskott-Aldrich disorder; blood coagulation disorder; Addison's disease;
KW autoimmune disorder; multiple sclerosis; systemic lupus erythematosus;
KW SLE; hyperproliferative disorder; gene therapy; neoplasm;
KW infectious disease; immunomodulatory; cytostatic; antimicrobial; ss.
XX
OS Homo sapiens.
XX
Key Location/Qualifiers
FH 44..826
CDS /tag= a
FT /product= "Human pituitary hormone, pituitrone"
FT sig_peptide 44..142
FT /tag= b
FT mat_peptide 143..823
FT /tag= c
FT /product= "Mature human pituitary hormone, pituitrone"
XX
WO20006778-A1.
XX
XX 09-NOV-2000.
XX
XX 27-APR-2000; 2000WO-US11211.
XX
XX 30-APR-1999; 99US-0131966.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, NI J;
XX
XX WPI: 2000-687547/67.
XX P-PSDB; AAY71959.
XX
XX Nucleic acids encoding a novel pituitary hormone (pituitrone), useful
XX for diagnosing, preventing and treating e.g. immune disorders,
XX hyperproliferative disorders and blood coagulation disorders -
XX
XX Claim 4; Fig 1; 277pp; English.
XX
XX The present sequence is a cDNA clone HKGDL36 encoding human pituitary
XX hormone, pituitrone. Pituitrone is highly expressed in pituitary
XX gland and is also expressed in brain tissues, spinal cord and kidney.
XX It may be used as antigens in the production of antibodies against
XX pituitrone and in assays to identify modulators. Pituitrone cDNAs are
XX also useful in gene therapy. Pituitrone may be used in the prevention,
XX treatment and diagnosis of diseases associated with inappropriate
XX pituitrone expression. It may be useful in treating disorders related
XX to reproductive and renal system, immune disorders, disorders of
XX haematopoietic cells (e.g., anaemia, Digeorge syndrome, ataxia
XX telangiectasia and Wiskott-Aldrich disorder), blood coagulation
XX disorders, autoimmune disorders (e.g. Addison's disease, multiple
XX sclerosis and systemic lupus erythematosus (SLE)), hyperproliferative
XX disorders (e.g. neoplasms of bone, liver and pancreas) and infectious
XX diseases.
XX
XX Sequence 1043 BP; 171 A; 391 C; 346 G; 135 T; 0 other;

	Query Match	91.08;	Score 891;	DB 21;	Length 1043;
	Best Local Similarity	99.9%;	pred. No. 0;		
	Matches 941;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
Qy	38	CCCCGGCTCGTGGGCGACATGCCGGGTCCCGCCTGCTCGGGGCGCGGGCGCGGGG	97		


```
Db 121 CGCGCCCTAAGCAGCGCTCTCCGCCCTTGCTGAGACTGCGGCTCTCTCGCGGCTTCGGG 180
QY 238 CGGTCAAGTGGCCCCAGAGTGAAGCGCGCGGGCGGTGAGAGCTGGCGCGGCGCTGGCG 297
Db 181 CGGTCAAGTGGCCCCAGAGTGAAGCGCGCGGGCGGTGAGAGCTGGCGCGGCGCTGGCG 240
QY 298 CATCTGCTGGAGGCGGACAGTCAAGAGGCGGCGGGCGGAGGCGGAGGAGCTGAGGAT 357
Db 241 CATCTGCTGGAGGCGGACAGTCAAGAGGCGGCGGGCGGAGGCGGAGGAGCTGAGGAT 300
QY 358 CAGCAGCGCGCGCTTCCTGGCGCGCTGCTGCGCTCTGGGGCGCGCGCGCAACTCTGAT 417
Db 301 CAGCAGCGCGCGCTTCCTGGCGCGCTGCTGCGCTCTGGGGCGCGCGCGCAACTCTGAT 360
QY 418 CCGGCTCTGGGCTTTGGACGAGACCGCGCGCTGCGAGCGAGCTCGCTTCGGCGCTG 477
Db 361 CCGGCTCTGGGCTTTGGACGAGACCGCGCGCTGCGAGCGAGCTCGCTTCGGCGCTG 420
QY 478 CTCCGGCGCGCGCTTGACCTTCGCGCGCTAGCAGCGCGCTTGCTCCGCGCGCGCTCC 537
Db 421 CTCCGGCGCGCGCTTGACCTTCGCGCGCTGCGAGCGCGCTGCGAGCGCGCTTCG 480
QY 538 GCCCGCGCGCTTCGACCGCGCGCGCTGACGAGCGCGCGCGCGCGCGCGGATGCT 597
Db 481 GCCCGCGCGCTTCGACCGCGCGCGCTGACGAGCGCGCGCGCGCGCGCGGATGCT 540
QY 598 GAGGAGCGAGCGGACGAGACCGCGCGCTGAGCGCGCGCGCGCGCGCGCTGCTGGGA 657
Db 541 GAGGAGCGAGCGGACGAGACCGCGCGCTGAGCGCGCGCGCGCGCGCGCTGCTGGGA 600
QY 658 CGGATTTCTGGGAAAGCGGAGCTCGAGGGGTGCGAGCGCGCGCGCGCGCGCGCT 717
Db 601 CGGATTTCTGGGAAAGCGGAGCTCGAGGGGTGCGAGCGCGCGCGCGCGCGCGCT 660
QY 718 GCCCGCGCGCGCTGAGCTGAGCTGCGCGCGCTGAGGCGCGCGCGCGCGCGCTGCT 777
Db 661 GCCCGCGCGCGCTGAGCTGAGCTGCGCGCGCTGAGGCGCGCGCGCGCGCGCTGCT 720
QY 778 CGTGTGAACGCGCTTAGAGACCGCGCGCGCGCGCGCTGAGCTGCGAGCGCGCGCG 837
Db 721 CGTGTGAACGCGCTTAGAGACCGCGCGCGCGCGCGCTGAGCTGCGAGCGCGCGCG 780

RESULT 8
AAZ98064
ID AAZ98064 standard; cDNA; 1050 BP.
XX
AC AAZ98064;
XX
DT 09-MAY-2000 (first entry)
XX
DE Human secreted protein encoding nucleotide sequence SEQ ID NO:58.
XX
KW Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
KW antiinflammatory; nontropic; neuroprotective; antiallergic; cancer;
KW tumour; neurodegenerative disorder; developmental abnormality; allergy;
KW foetal deficiency; blood disorder; immune system disorder; arthritis;
KW autoimmune disease; hepatic disease; renal disease; inflammation;
KW Alzheimer's disease; behavioural disorder; schizophrenia; osteoporosis;
KW infection; AIDS; spinal cord injury; transplant rejection; diabetes;
KW asthma; sepsis; acne; psoriasis; cardiovascular disorder;
KW reproductive disorder; gastrointestinal disorder; respiratory disorder;
KW metabolic disorder; food additive; preservative; ss.
XX
OS Homo sapiens.
XX
PN WO200004140-A1.
XX
PD 27-JAN-2000.
XX
PF 14-JUL-1999; 99WO-US15849.
XX
PR 15-JUL-1998; 98US-0092921.
```

```
PR 15-JUL-1998; 98US-0092922.
XX 15-JUL-1998; 98US-0092956.
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Komatsoulis G, Duan RD, Rosen CA, Moore PA, Shi Y;
PI Lafleur DM, Ebner R, Olsen HS, Brewer LA, Florence KA, Young PE;
PI Mucenski M, Endress GA, Soppet DR;
XX
DR WPI: 2000-161128/14.
DR P-PSDB; AAY87111.
XX
PT New isolated human genes, useful for diagnosis and treatment of, e.g.
PT cancers, neurological or blood disorders
PS Claim 1; Page 339; 494pp; English.
XX
CC The polynucleotide sequences given in AAZ98017 to AAZ98108 encode the
CC human secreted proteins given in AAY87064 to AAY87223. Human secreted
CC protein can have activities based on the tissues and cells the genes are
CC expressed in. Examples of activities include: cytostatic;
CC immunosuppressive; antiinflammatory; nontropic; neuroprotective; and
CC antiallergic. The polynucleotides and their corresponding secreted
CC polypeptides are useful for preventing, treating or ameliorating medical
CC conditions, e.g. by protein or gene therapy. Also pathological conditions
CC can be diagnosed by determining the amount of the new polypeptides in a
CC sample or by determining the presence of mutations in the new
CC polynucleotides. Human secreted protein s and their polynucleotides can
CC be used for developing products for the diagnosis or treatment of cancer,
CC tumours, neurodegenerative disorders, developmental abnormalities and
CC foetal deficiencies, blood disorders, diseases of the immune system,
CC autoimmune diseases, hepatic and renal disease, inflammation,
CC allergies, Alzheimer's disease, behavioural disorders, schizophrenia,
CC osteoporosis, arthritis, infections, AIDS, spinal cord injuries,
CC transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,
CC cardiovascular disorders, reproductive disorders, gastrointestinal
CC disorders, respiratory disorders and metabolic disorders. The
CC proteins or polynucleotides can also be used as food additives or
CC preservatives. The proteins are also useful for identifying their
CC binding partners. AAZ98008 to AAZ98016 and AAY87063 are sequence used in
CC the exemplification of the present invention.
XX
SQ Sequence 1050 BP; 172 A; 398 C; 344 G; 136 T; 0 other;

Query Match 59.6%; Score 583; DB 21; Length 1050;
Best Local Similarity 99.8%; Pred. No. 3.1e-250;
Matches 633; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 346 GAGGCTGAGGATCAGCAGGCGCGGCTCTGCGCAGCTGCTGCGGCGCGCC 405
Db 339 GAGGCTGAGGATCAGCAGGCGCGGCTCTGCGCAGCTGCTGCGGCGCGCC 398
QY 406 CGCAACTCTCATCGGCTCTGGGCTTGAGCAGACACCCGACGCGCTGAGCGCAGCTC 465
Db 399 CGCAACTCTCATCGGCTCTGGGCTTGAGCAGACACCCGACGCGCTGAGCGCAGCTC 458
QY 466 GCTCGGCGCTGCTGCTCCGCGCGCGCTTGACCTGCGCGCGCTAGCAGCCAGCTTGTCCCC 525
Db 459 GCTCGGCGCTGCTGCTCCGCGCGCGCTTGACCTGCGCGCGCTAGCAGCCAGCTTGTCCCC 518
QY 526 GCGGCGCTCCCGCGCGCGGCTGCGACCGCGCGCGCGCGCGCTTACGACAGCGCGCGCG 585
Db 519 GCGGCGCTCCCGCGCGCGGCTGCGACCGCGCGCGCGCGCTTACGACAGCGCGCGCG 578
QY 586 GGCGCGGATCTGAGGAGGCGGAGCAGACACCCGACGCTGGACCCCGAGCTTTCAGG 645
Db 579 GGCGCGGATCTGAGGAGGCGGAGCAGACACCCGACGCTGGACCCCGAGCTTTCAGG 638
QY 646 TACTTGTCTGGGAGCGATTCTTGGGGAAAGCGCGGACTTCCGAGGGGGTGGCAGCCCCCGCG 705
Db 639 TACTTGTCTGGGAGCGATTCTTGGGGAAAGCGCGGACTTCCGAGGGGGTGGCAGCCCCCGCG 698
QY 706 CGCTCGCGCGCTGCGCGCGGACGAGATGTGGGCTCTGAGCTGCCCGCTGAGGGGGTGTCTG 765
```

```
Db 699 CGCCTCGCGCTGCGCGGACACAGATGGGCTCTGAGCTGCCCTGAGGGCGTGTG 758
QY 766 GGGGCGCTGCTGGTGTGAACGCTTAGAGACCCCGCGCGCCAGGTGCTGCACGCCGC 825
Db 759 GGGGCGCTGCTGGTGTGAACGCTTAGAGACCCCGCGCGCCAGGTGCTGCACGCCGC 818
QY 826 CTCTTGGCCACCTGAGCACTGCGCGGATCCGCTGACCTGGGACCCAGAGTGGCCCGC 885
Db 819 CTCTTGGCCACCTGAGCACTGCGCGGATCCGCTGACCTGGGACCCAGAGTGGCCCGC 878
QY 886 CCATCCCGCCACAGCACTGCTCCCGCGCAGCAGCTCCAGAGCACTTACCCCGCGCAGC 945
Db 879 CCATCCCGCCACAGCACTGCTCCCGCGCAGCAGCTCCAGAGCACTTACCCCGCGCAGC 938
QY 946 CAGCCCTCTACCCGAGGATCCCTACCCCTGCG 979
Db 939 CAGCCCTCTACCCGAGGATCCCTACCCCTGCG 972

RESULT 9
AADI1677
ID AADI1677 standard; cDNA; 1050 BP.
AC AADI1677;
XX
XX
XX 24-SEP-2001 (first entry)
DE
DE Human secreted protein-encoding gene 48 cDNA clone HKGDL36, SEQ ID NO:58.
KW Human; secreted protein; proliferative disorder; cancer; tumour; asthma;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disease; AIDS; autoimmune disease; rheumatoid arthritis;
KW Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;
KW psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;
KW inflammation; neurological disorder; Alzheimer's disease; food additive;
KW angogenic disorder; kidney disorder; gastrointestinal disorder; allergy;
KW pregnancy-related disorder; endocrine disorder; infection; wound healing;
KW cell culture; chemotaxis; vulnery; binding partner identification;
KW gene therapy; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 55..501
XX /tag= a
XX /product= "Human secreted protein precursor"
XX sig_peptide 55..153
XX /tag= b
XX mat_peptide 154..498
XX /tag= c
XX /product= "Mature human secreted protein"
XX
XX WO200151504-A1.
XX
XX 19-JUL-2001.
XX
XX 12-JAN-2001; 2001WO-US00911.
XX
XX 13-JAN-2000; 2000US-0482273.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Komatsoulis CA, Duan DR, Rosen CA, Moore PA, Shi Y;
XX Lafleur DW, Olsen HS, Brewer LA, Florence KA, Young PE, Soppet DR;
XX Endress GA, Muscenski M, Ebner R;
XX
XX WPI; 2001-425865/45.
XX P-PSDB; AAE06088.
XX
XX Isolated nucleic acid molecule encoding a human secreted protein is
XX used in preventing, treating or ameliorating a medical condition -
```

```
PS
XX
XX AADI1630-AADI1721 represent cDNAs corresponding to 71 human secreted
XX protein genes, and AAE06041-AAE06132 represent the proteins they encode.
XX AAE06133-AAE06205 represent human secreted protein fragments.
XX The secreted proteins and their genes are useful for preventing, treating
XX or ameliorating medical conditions, e.g., by protein or gene therapy.
XX Pathological conditions can be diagnosed by determining the amount of the
XX new protein in a sample or by determining the presence of mutations in
XX the new genes. Specific uses are described for each of the 71 genes,
XX based on the tissues in which they are most highly expressed, and include
XX developing products for the diagnosis or treatment of proliferative
XX disorders, cancer, tumours, foetal and developmental abnormalities,
XX haematopoietic disorders, diseases of the immune system, AIDS, autoimmune
XX diseases (e.g., rheumatoid arthritis), inflammation, allergies,
XX neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),
XX cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,
XX psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,
XX angogenic disorders, kidney disorders, gastrointestinal disorders,
XX pregnancy-related disorders, endocrine disorders, and infectious. The
XX proteins can also be used to aid wound healing and epithelial cell
XX proliferation, to prevent skin aging due to sunburn, to maintain organs
XX before transplantation, for supporting cell culture of primary tissues,
XX to regenerate tissues, to identify their cognate ligands or binding
XX partners, and in chemotaxis, and can be used as a food additive or
XX preservative to modify storage properties. Antibodies specific for a
XX protein of the invention can be used in alleviating symptoms associated
XX with the disorders mentioned above, and in diagnostic immunoassays e.g.,
XX radioimmunoassay or enzyme linked immunosorbent assay (ELISA).
XX The present sequence represents a human secreted protein-encoding cDNA of
XX the invention.
SQ Sequence 1050 BP; 172 A; 398 C; 344 G; 136 T; 0 other;
Query Match 59.6%; Score 583; DB 22; Length 1050;
Best Local Similarity 99.8%; Pred. No. 3.1e-250;
Matches 633; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 346 GAGGCTGAGGATCAGCAGCGCGCTCTGGCGCAGCTGCTGCGCGCGGCGCC 405
Db 339 GAGGCTGAGGATCAGCAGCGCGCTCTGGCGCAGCTGCTGCGCGCGGCGCC 398
QY 406 CGCAACTCTGATCCGGCTTGGGCTTGGAGCAGACCCCGACGCGCTGAGCGCAGCTC 465
Db 399 CGCAACTCTGATCCGGCTTGGGCTTGGAGCAGACCCCGACGCGCTGAGCGCAGCTC 458
QY 466 GCTCGCGCTGCTGCTCGCGCGCGCTTGGAGCAGCTGCTGCGCGCTGCTGCGCGCG 525
Db 459 GCTCGCGCTGCTGCTCGCGCGCGCTTGGAGCAGCTGCTGCGCGCTGCTGCGCGCG 518
QY 526 GCGCCCGTCCCGCGCGCGCTTGGAGCAGCTGCTGCGCGCGCTTGGAGCAGCTGCTGCG 585
Db 519 GCGCCCGTCCCGCGCGCGCTTGGAGCAGCTGCTGCGCGCGCTTGGAGCAGCTGCTGCG 578
QY 586 GGGCCGGATGCTGAGGAGCAGCGCAGCAGACACCCCGACGCTGAGCCCGAGCTGTTGAGG 645
Db 579 GGGCCGGATGCTGAGGAGCAGCGCAGCAGACACCCCGACGCTGAGCCCGAGCTGTTGAGG 638
QY 646 TACTTCTGGGACGGATTTCTTGGGGAAGCGCGAGCTCCGAGGGGGTGGCAGCCCGCGC 705
Db 639 TACTTCTGGGACGGATTTCTTGGGGAAGCGCGAGCTCCGAGGGGGTGGCAGCCCGCGC 698
QY 706 CGCCTCGCGCGTCCCGCGCGCGCTTGGAGCAGCTGCTGAGCTGCGCCCGCTGAGGCGCTGCTG 765
Db 699 CGCCTCGCGCGTCCCGCGCGCGCTTGGAGCAGCTGCTGAGCTGCGCCCGCTGAGGCGCTGCTG 758
QY 766 GGGGCGCTGCTGCTGTGAACGCTTAGAGACCCCGCGCGCCAGGTGCTGCACGCCGC 825
Db 759 GGGGCGCTGCTGCTGTGAACGCTTAGAGACCCCGCGCGCCAGGTGCTGCACGCCGC 818
QY 826 CTCTTGGCCACCTGAGCACTGCGCGGATCCGCTGACCTGGGACCCAGAGTGGCCCGC 885
Db 819 CTCTTGGCCACCTGAGCACTGCGCGGATCCGCTGACCTGGGACCCAGAGTGGCCCGC 878
```

QY 886 CCATCCGCGCCACAGGACTGCTCCCGCCAGCAGCTCCAGAGCAACTTACCCCGGCAGC 945
|||||
Db 879 CCATCCGCGCCACAGGACTGCTCCCGCCAGCAGCTCCAGAGCAACTTACCCCGGCAGC 938
QY 946 CAGCCCTCTACCCGAGGATCCCTACCCCTGGC 979
|||||
Db 939 CAGCCCTCTACCCGAGGATCCCTACCCCTGGC 972

RESULT 10
ABK69773
ID ABK69773 standard; cDNA; 1050 BP.
XX
AC ABK69773;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human secreted protein gene 48.
XX
KW Human; ss; gene; secreted protein; gene therapy; immunosuppressive;
KW antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
KW vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
KW virucide; fungicide; ophthalmological; autoimmune disease; neoplasm;
KW rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;
KW cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;
KW angiogenesis; nervous system disorder; Alzheimer's disease; infection;
KW ocular disorder; corneal infection; wound healing; skin aging;
KW epithelial cell proliferation; food additive.
XX
OS Homo sapiens.
XX
PN WO200226931-A2.
XX
PD 04-APR-2002.
XX
PF 24-SEP-2001; 2001WO-US29871.
XX
PR 25-SEP-2000; 2000US-234925P.
PR 12-JAN-2001; 2001WO-US00911.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Komatsoulis G, Duan DR, Rosen CA, Moore PA, Shi Y;
PI Lafleur DW, Olsen H, Brewer LA, Florence KA, Young PE, Soppet DR;
PI Endress CA, Mucenski M, Ebner R;
XX
DR WPI: 2002-362489/39.
DR P-PSDB: ABG33910, ABG34000.
XX
PT Novel 71 isolated secreted polypeptides and polynucleotides encoding
PT the polypeptides, useful for treating Huntington's disease, sepsis,
PT meningitis, thrombocytopenia, haemolytic anaemia, rheumatoid arthritis,
PT asthma
XX
PS Claim 1; Page 1189; 1478pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (or its
CC fragment, homologue complement or allelic variant) encoding a human
CC secreted protein (and its fragment, domain, epitope, variant, secreted
CC form and species variant). Also included are a recombinant vector
CC comprising the nucleic acid, a recombinant host cell comprising the
CC vector, an antibody against the secreted protein, a recombinant host cell
CC that expresses the secreted protein and a method of identifying a binding
CC partner of the secreted protein. The nucleic acid and protein are used to
CC prevent, diagnose, treat or ameliorate a medical condition in e.g.
CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep
CC for example autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi and
CC ocular disorders e.g. corneal infection. Many other diseases and

CC disorders are listed in the specification. The polypeptides can also be
CC used to aid wound healing an epithelial cell proliferation, to prevent
CC skin aging due to sunburn, to maintain organs before transplantation, for
CC supporting cell culture of primary tissues, to regenerate tissues and in
CC chemotaxis. The polypeptides can also be used as a food additive or
CC preservative to increase or decrease storage capabilities. The present
CC sequence encodes a novel human secreted protein of the invention.
XX

SQ Sequence 1050 BP; 172 A; 398 C; 344 G; 136 T; 0 other;

Query Match 59.68; Score 583; DB 24; Length 1050;
Best Local Similarity 99.8%; Pred. No. 3.1e-250;
Matches 633; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 346 GAGGCTGAGGATCAGCAGCGCGCTCTGGCGCAGCTGCTGGCGCTCTGGCGCGCCCGCC 405
|||||
Db 339 GAGGCTGAGGATCAGCAGCGCGCTCTGGCGCAGCTGCTGGCGCTCTGGCGCGCCCGCC 398
|||||
QY 406 CGCAACTCTGATCGGCTCTGGGCTTGAGACGACGACCCCGCGCTTCAGCGAGCTC 465
|||||
Db 399 CGCAACTCTGATCGGCTCTGGGCTTGAGACGACGACCCCGCGCTTCAGCGAGCTC 458
|||||
QY 466 GCTCGGCTCTGCTCCGGCGCGCGCTTGACCTCGCGCGCTTCAGCGAGCTC 525
|||||
Db 459 GCTCGGCTCTGCTCCGGCGCGCGCTTGACCTCGCGCGCTTCAGCGAGCTC 518
|||||
QY 526 GCGCCGCTCCCGCGCGCGCGCTCCGACCCCGCGCGCGCTTCAGCGAGCTC 585
|||||
Db 519 GCGCCGCTCCCGCGCGCGCGCTCCGACCCCGCGCGCGCTTCAGCGAGCTC 578
|||||
QY 586 GGCCCGGATGCTGAGGAGGCGAGCGAGACACCCCGAGCTGAGCTGTTGAGG 645
|||||
Db 579 GGCCCGGATGCTGAGGAGGCGAGCGAGACACCCCGAGCTGAGCTGTTGAGG 638
|||||
QY 646 TACTTGTGGACGGATTCTTGGGGAAGCGCGACATCCGAGGGGTGGCAGCCCGCGC 705
|||||
Db 639 TACTTGTGGACGGATTCTTGGGGAAGCGCGACATCCGAGGGGTGGCAGCCCGCGC 698
|||||
QY 706 CGCCTCGCGCTGCGCGCGCGCGCGCTGAGCTGCGCGCTGAGCTGCGCGCTGAGCTG 765
|||||
Db 699 CGCCTCGCGCTGCGCGCGCGCGCGCTGAGCTGCGCGCTGAGCTGCGCGCTGAGCTG 758
|||||
QY 766 GGGCGCTGCTGCTGTGAACGCTTAGAGACCCCGCGCGCGCGCTGAGCTGCGCGCTG 825
|||||
Db 759 GGGCGCTGCTGCTGTGAACGCTTAGAGACCCCGCGCGCGCGCTGAGCTGCGCGCTG 818
|||||
QY 826 CTCTTGCACCTCAGCAGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTG 885
|||||
Db 819 CTCTTGCACCTCAGCAGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTG 878
|||||
QY 886 CCATCCGCGCCACAGGACTGCTCCCGCCAGCAGCTCCAGAGCAACTTACCCCGGCAGC 945
|||||
Db 879 CCATCCGCGCCACAGGACTGCTCCCGCCAGCAGCTCCAGAGCAACTTACCCCGGCAGC 938
|||||
QY 946 CAGCCCTCTACCCGAGGATCCCTACCCCTGGC 979
|||||
Db 939 CAGCCCTCTACCCGAGGATCCCTACCCCTGGC 972
|||||

RESULT 11

AA158162

ID AA158162 standard; cDNA; 919 BP.

XX

AC AA158162;

XX

DT 22-OCT-2001 (first entry)

XX Human polynucleotide SEQ ID NO 365.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;

XX peripheral nervous system; neuropathy; central nervous system; CNS;

XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;

XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

PS Example 2; Page 1404-1406; 1478pp; English.

XX The invention relates to an isolated nucleic acid molecule (or its
CC fragment, homologue complement or allelic variant) encoding a human
CC secreted protein (and its fragment, domain, epitope, variant, secreted
CC form and species variant). Also included are a recombinant vector
CC comprising the nucleic acid, a recombinant host cell comprising the
CC vector, an antibody against the secreted protein, a recombinant host cell
CC that expresses the secreted protein and a method of identifying a binding
CC partner of the secreted protein. The nucleic acid and protein are used to
CC prevent, diagnose, treat or ameliorate a medical condition in e.g.
CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep
CC for example autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi and
CC ocular disorders e.g. corneal infection. Many other diseases and
CC disorders are listed in the specification. The polypeptides can also be
CC used to aid wound healing an epithelial cell proliferation, to prevent
CC skin aging due to sunburn, to maintain organs before transplantation, for
CC supporting cell culture of primary tissues, to regenerate tissues and in
CC chemotaxis. The polypeptides can also be used as a food additive or
CC preservative to increase or decrease storage capabilities. The present
CC sequence represents a ds DNA fragment of the gene for a novel human
CC secreted protein of the invention.

XX
SQ Sequence 4503 BP; 783 A; 1181 C; 1314 G; 1225 T; 0 other;

Query Match 38.3%; Score 375; DB 24; Length 4503;
Best Local Similarity 99.6%; Pred. No. 1.5e-137;
Matches 475; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 170 AGGAGCCCGCGGCTAGACGAGCGTCTCGCGCTTGGCTGAGACTGGCGCTCTCGCC 229
Db 3254 AGGAGCCCGCGGCTAGACGAGCGTCTCGCGCTTGGCTGAGACTGGCGCTCTCGCC 3313

Qy 230 GCTTCGCGGCTAGTGGCCCGAGGTAGGCGGCGGCGGCGGTGAGAGTGGCGCGG 289
Db 3314 GCTTCGCGGCTAGTGGCCCGAGGTAGGCGGCGGCGGCGGTGAGAGTGGCGCGG 3373

Qy 290 CGCTGGCGCATCTCTGGAGCGCGAAGCTAGGAGCGGCGGCGGCGGCGGCGGAGG 349
Db 3374 CGCTGGCGCATCTCTGGAGCGCGAAGCTAGGAGCGGCGGCGGCGGCGGCGGAGG 3433

Qy 350 CTGAGGATCAGCAGCGCGCTCTCTGGCGCAGCTGCTGCGGCTCTGGGCGCGCCCGCA 409
Db 3434 CTGAGGATCAGCAGCGCGCTCTCTGGCGCAGCTGCTGCGGCTCTGGGCGCGCCCGCA 3493

Qy 410 ACTGTATCCGGCTCTGGGCTTGGACGACGACACCGCGAGCGGCTGCGAGCTCGGTC 469
Db 3494 ACTGTATCCGGCTCTGGGCTTGGACGACGACACCGCGAGCGGCTGCGAGCTCGGTC 3553

Qy 470 GCGTCTGCTCCGCGCGGCTTACCCCTGCGGCTAGCAGCCAGCTTGTCCCGCGC 529
Db 3554 GCGTCTGCTCCGCGCGGCTTACCCCTGCGGCTAGCAGCCAGCTTGTCCCGCGC 3613

Qy 530 CCGTCCCGCGCGGCGGCTCCGACCGCGGCGCGGCTTACGACGAGCGGCGCGGCGCC 589
Db 3614 CCGTCCCGCGCGGCGGCTCCGACCGCGGCGCGGCTTACGACGAGCGGCGCGGCGCC 3673

Qy 590 CGGATGTGAGGAGCGGCGGACGAGACACCGGACGTGGACCCCGAGCTGTGTAGGT 646
Db 3674 CGGATGTGAGGAGCGGCGGACGAGACACCGGACGTGGACCCCGAGCTGTGTAGGT 3730

RESULT 13
ABN95203/c
ID ABN95203 standard; DNA; 301 BP.
XX
XX ABN95203;
XX
XX 13-AUG-2002 (first entry)

XX Gene #1701 used to diagnose liver cancer.

DE Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
XX metastatic liver tumour; cytostatic; expression profile; disease state;
KW disease progression; drug toxicity; drug efficacy; drug metabolism.
XX Homo sapiens.
XX WO200229103-A2.
XX 11-APR-2002.
XX 02-OCT-2001; 2001WO-US30589.
XX 02-OCT-2000; 2000US-237054P.
XX (GENE-) GENE LOGIC INC.
XX Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX WPI; 2002-426119/45.
XX Diagnosing and detecting the progression of liver cancer,
PT hepatocellular carcinoma or metastatic liver tumor in a patient,
PT involves detecting the level of expression of two or more genes in a
PT liver tissue sample
XX
XX Claim 1; SEQ ID NO 1701; 298pp; English.

XX The invention relates to a novel method for diagnosing and detecting the
CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC tumour in a patient, and differentiating metastatic liver cancer from
CC hepatocellular carcinoma in a patient, involving detecting the level of
CC expression of two or more genes represented in ABN93503-ABN97455 in a
CC tissue sample. The method of the invention has hepatotropic, and
CC cystostatic activity. The method is useful for diagnosing and detecting
CC the progression of liver cancer, hepatocellular carcinoma and metastatic
CC liver carcinoma in a patient. The method is useful for identifying
CC expression profiles which serve as useful diagnostic markers as well as
CC markers that can be used to monitor disease states, disease progression,
CC drug toxicity, drug efficacy and drug metabolism.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 301 BP; 41 A; 79 C; 131 G; 48 T; 2 other;

Query Match 23.1%; Score 226; DB 24; Length 301;
Best Local Similarity 99.6%; Pred. No. 5e-91;
Matches 276; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 703 CGCGCGCTCCGCGCTGCGCGGCGGCGGCGGCTGAGCTGCGGCTGAGGCGGCTG 762
Db 301 CGCGCGCTCCGCGCTGCGCGGCGGCGGCGGCTGAGCTGCGGCTGAGGCGGCTG 242

Qy 763 CTGCGGCGGCTGCTGCGCTGTGAAACGCGCTAGACACCGCGGCGGCGGCTGCGTGCACGC 822
Db 241 CTGCGGCGGCTGCTGCGCTGTGAAACGCGCTAGACACCGCGGCGGCGGCTGCGTGCACGC 182

Qy 823 GCGCTCTTGGCACCGCTGAGCAGCTGCGGCGGATCCCGTGCACCGCTGGGACCCAGAGTGC 882
Db 181 GCGCTCTTGGCACCGCTGAGCAGCTGCGGCGGATCCCGTGCACCGCTGGGACCCAGAGTGC 122

Qy 883 CGCGCATCCCGCGGCGGCGGCTGCGGCGGCGGCGGCGGCTGAGGCGGCGGCGGCGG 942
Db 121 CGCGCATCCCGCGGCGGCGGCTGCGGCGGCGGCGGCGGCTGAGGCGGCGGCGGCGG 62

Qy 943 AGCGAGCGCTCTACCGCGGAGGATCCCTACCGCGGCGGCGGCGGCGGCGGCGGCGG 979
Db 61 AGCGAGCGCTCTACCGCGGAGGATCCCTACCGCGGCGGCGGCGGCGGCGGCGGCGG

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2003, 14:14:10 ; Search time 87 Seconds
(without alignments)
615.773 Million cell updates/sec

Title: US-09-803-589-6

Perfect score: 1319

Sequence: 1 MAGSPLLMGPRAGGVGLLV.....RVKRLTPAPQVARRLLPP 260

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organalle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1319	100.0	260	4	Q9UHG2	Q9uhg2 homo sapien
2	1102	83.5	260	11	Q9QXU9	Q9qxu9 rattus norv
3	1082	82.0	258	11	Q91W26	Q91w26 mus musculu
4	1076	81.6	258	11	Q9QXV0	Q9qxv0 mus musculu
5	490.5	37.2	187	11	Q9ESU4	Q9esu4 mus musculu
6	146	11.1	913	16	Q8UJ51	Q8uj51 agrobacteri
7	136.5	10.3	388	4	Q9NY42	Q9ny42 homo sapien
8	136.5	10.3	441	4	Q75685	Q75685 homo sapien
9	134	10.2	553	16	Q9RSN4	Q9rsn4 deinococcus
10	132.5	10.0	612	2	Q9Z614	Q9z614 streptomyce
11	131	9.9	492	10	Q9RV89	Q9rv89 oryza sativ
12	131	9.9	1290	5	Q9NEA7	Q9nea7 leishmania
13	130.5	9.9	589	16	Q9A718	Q9a718 caulobacter
14	130	9.9	350	2	Q9EU24	Q9eu24 corynebacte
15	129.5	9.8	544	11	Q9VCG2	Q9vcg2 mus musculu
16	129	9.8	660	16	Q9RDI3	Q9rdi3 streptomyce

17	127.5	9.7	460	5	Q9GZF7	Q9gzf7 caenorhabdi
18	125.5	9.5	431	16	Q9L0J8	Q9l0j8 streptomyce
19	125.5	9.5	816	4	Q9G651	Q9gd51 homo sapien
20	125.5	9.5	1340	16	Q91LH8	Q91lh8 streptomyce
21	124.5	9.4	4809	2	Q93HH0	Q93hh0 streptomyce
22	123.5	9.4	700	2	Q34003	Q34003 rhodobacter
23	123	9.3	801	5	Q23635	Q23635 caenorhabdi
24	121.5	9.2	590	16	Q9S2Q5	Q9s2q5 streptomyce
25	121	9.2	772	16	Q9AD96	Q9ad96 streptomyce
26	121	9.2	791	11	Q9Z1P7	Q9z1p7 mus musculu
27	120.5	9.1	817	4	Q96SB3	Q96sb3 homo sapien
28	119	9.0	421	5	Q9VGC1	Q9vgc1 drosophila
29	119	9.0	469	12	Q69270	Q69270 equine herp
30	119	9.0	934	16	Q9K452	Q9k452 streptomyce
31	118.5	9.0	667	16	Q9RJY5	Q9rjy5 streptomyce
32	118.5	9.0	1641	5	Q9GRZ3	Q9grz3 caenorhabdi
33	118	8.9	238	11	Q9JLJ0	Q9jlj0 mus musculu
34	118	8.9	1240	12	Q9DWH8	Q9dwh8 rat cytomeg
35	118	8.9	1430	11	Q8VHK2	Q8vnh2 rattus norv
36	117.5	8.9	327	11	Q99JK6	Q99jk6 mus musculu
37	117.5	8.9	615	4	Q9UDW8	Q9udw8 homo sapien
38	117.5	8.9	941	16	Q9ZIW5	Q9ziw5 streptomyce
39	117	8.9	406	2	Q8VPM8	Q8vpm8 micrococcus
40	117	8.9	2157	11	Q9Z1R1	Q9z1r1 mus musculu
41	116.5	8.8	1198	11	Q9JKA7	Q9jka7 rattus norv
42	116.5	8.8	1677	11	Q70373	Q70373 mus musculu
43	116.5	8.8	2307	2	Q9AG79	Q9ag79 streptomyce
44	116.5	8.8	6146	2	Q93HJ5	Q93hj5 streptomyce
45	116	8.8	710	4	Q9H0J3	Q9h0j3 homo sapien

ALIGNMENTS

RESULT 1

Q9UHG2 ID Q9UHG2 PRELIMINARY; PRT; 260 AA.

AC Q9UHG2;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PROSNAAS precursor (Granin-like neuroendocrine peptide precursor).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Fricker L., McKinzie A.A., Sun J., Curran E., Qian Y., Yan L.,
RA Patterson S.D., Courchesne P.L., Richards B., Levin N., Mzhavia N.,
RA Devi L.A., Douglass J.;
RT *Identification and characterization of prosnaas: a granin-like
RT neuroendocrine peptide precursor that inhibits prohormone
RT processing*;
RL J. Neurosci. 20:0-0(2000).
RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=UTERUS;

RA Strausberg R.;

RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF181562; AAF22643.1; -

DR EMBL; BC002851; AAH02851.1; -

KW Signal.

FT SIGNAL 1 33 POTENTIAL.

SQ SEQUENCE 260 AA; 27372 MW; FF8E2722784B7A5C CRC64;

Query Match 100.0%; Score 1319; DB 4; Length 260;

Best Local Similarity 100.0%; Pred. NO. 3.8e-74;

Matches 260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGSPLLMGPRAGGVGLLVLLLLGLFRPPALCARPVKEPRGLSAASPPLAETGAPRRR 60

|||||

DB 1 MAGSPLLMGPRAGGVGLLVLLLLGLFRPPALCARPVKEPRGLSAASPPLAETGAPRRR 60

```
QY 61 RSVPRGEAAGAVQELARALAHLLERQERARAEAEQEDQARVLAQLLRVWGAPRNSD 120
DB 61 RSVPRGEAAGAVQELARALAHLLERQERARAEAEQEDQARVLAQLLRVWGAPRNSD 120
QY 121 PALGLDDDDPAPAAQALARALLRRLDPAALAAQLVPAPVPAALRPPVYDDGPGDPA 180
DB 121 PALGLDDDDPAPAAQALARALLRRLDPAALAAQLVPAPVPAALRPPVYDDGPGDPA 180
QY 181 EAGDETDPDPELLRYLLGRILAGSADSEGVAAAPRRRLRAADHDVDSGSELPEGVIGALL 240
DB 181 EAGDETDPDPELLRYLLGRILAGSADSEGVAAAPRRRLRAADHDVDSGSELPEGVIGALL 240
QY 241 RVKRLTAPQVPAARRLLPP 260
DB 241 RVKRLTAPQVPAARRLLPP 260

RESULT 2
QY 61 RSVPRGEAAGAVQELARALAHLLERQERARAEAEQEDQARVLAQLLRVWGAPRNSD 120
DB 61 RSVPRGEAAGAVQELARALAHLLERQERARAEAEQEDQARVLAQLLRVWGAPRNSD 120
QY 121 PALGLDDDDPAPAAQALARALLRRLDPAALAAQLVPAPVPAALRPPVYDDGPGDPA 180
DB 121 PALGLDDDDPAPAAQALARALLRRLDPAALAAQLVPAPVPAALRPPVYDDGPGDPA 180
QY 181 EAGDETDPDPELLRYLLGRILAGSADSEGVAAAPRRRLRAADHDVDSGSELPEGVIGALL 240
DB 181 EAGDETDPDPELLRYLLGRILAGSADSEGVAAAPRRRLRAADHDVDSGSELPEGVIGALL 240
QY 241 RVKRLTAPQVPAARRLLPP 260
DB 241 RVKRLTAPQVPAARRLLPP 260

Query Match 82.0%; Score 1082; DB 11; Length 258;
Best Local Similarity 83.5%; Pred. No. 1.5e-59;
Matches 217; Conservative 10; Mismatches 31; Indels 2; Gaps 1;

AC Q91W26; PRELIMINARY; PRT; 260 AA.
ID Q90XU9;
AC Q90XU9;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE PROSAAS precursor.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20098938; PubMed=10632593;
RA Fricker L., McKinzie A.A., Sun J., Curran E., Qian Y., Yan L.,
RA Patterson S.D., Courchesne P.L., Richards B., Levin N., Mzhavia N.,
RA Devi L.A., Douglass J.
RT "Identification and characterization of proSAAS, a granin-like
RT neuroendocrine peptide precursor that inhibits prohormone
RT processing."
RL J. Neurosci. 20:639-648(2000).
DR EMBL; AF181561; AAF22642.1; -.
KW Signal.
FT SIGNAL.
SQ SEQUENCE 260 AA; 27414 MW; C84F688BDEB5313E CRC64;

Query Match 83.5%; Score 1102; DB 11; Length 260;
Best Local Similarity 84.6%; Pred. No. 8.6e-61;
Matches 220; Conservative 8; Mismatches 32; Indels 0; Gaps 0;

QY 1 MAGSPLLWGPRAGGVGLLVLLGLFRPPPPALCARPVKEPRGLSAASPLAETGAPRRFR 60
DB 1 MAGSPLLWGPRAGGVGLLVLLGLFRPPPPALCARPVKEPRGLSAASPLAETGAPRRFR 60
QY 61 RSVPRGEAAGAVQELARALAHLLERQERARAEAEQEDQARVLAQLLRVWGAPRNSD 120
DB 61 RSVPRGEAAGAVQELARALAHLLERQERARAEAEQEDQARVLAQLLRVWGAPRNSD 120
QY 121 PALGLDDDDPAPAAQALARALLRRLDPAALAAQLVPAPVPAALRPPVYDDGPGDPA 180
DB 121 PALGLDDDDPAPAAQALARALLRRLDPAALAAQLVPAPVPAALRPPVYDDGPGDPA 180
QY 181 EAGDETDPDPELLRYLLGRILAGSADSEGVAAAPRRRLRAADHDVDSGSELPEGVIGALL 240
DB 181 EAGDETDPDPELLRYLLGRILAGSADSEGVAAAPRRRLRAADHDVDSGSELPEGVIGALL 240
QY 241 RVKRLTAPQVPAARRLLPP 260
DB 241 RVKRLTAPQVPAARRLLPP 260

Query Match 82.0%; Score 1082; DB 11; Length 258;
Best Local Similarity 83.5%; Pred. No. 1.5e-59;
Matches 217; Conservative 10; Mismatches 31; Indels 2; Gaps 1;

AC Q91W26; PRELIMINARY; PRT; 258 AA.
ID Q90XV0;
AC Q90XV0;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE PROSAAS precursor.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20098938; PubMed=10632593;
RA Fricker L., McKinzie A.A., Sun J., Curran E., Qian Y., Yan L.,
RA Patterson S.D., Courchesne P.L., Richards B., Levin N., Mzhavia N.,
RA Devi L.A., Douglass J.
RT "Identification and characterization of proSAAS, a granin-like
RT neuroendocrine peptide precursor that inhibits prohormone
RT processing."
RL J. Neurosci. 20:639-648(2000).
DR EMBL; AF181560; AAF22641.1; -.
KW Signal.
FT SIGNAL.
SQ SEQUENCE 258 AA; 27285 MW; 07452A460E868CF6 CRC64;
```

ID	Q8UJ51	PRELIMINARY;	PRT;	913 AA.
AC	Q8UJ51:			
DT	01-JUN-2002 (Tremblrel. 21, Created)			
DT	01-JUN-2002 (Tremblrel. 21, Last sequence update)			
DT	01-JUN-2002 (Tremblrel. 21, Last annotation update)			
GN	Translation initiation factor IF-2.			
DE	INFB OR ATU0087 OR AGR.C131.			
OS	Agrobacterium tumefaciens (strain C58 / ATCC 33970).			
OC	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;			
OC	Rhizobiaceae; Rhizobium.			
ON	NCBI_TaxID=176299;			
OX	[1]			
RP	SEQUENCE FROM N.A.			
RP	MEDLINE=21608550; PubMed=11743193;			
RA	Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,			
RA	Okura Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,			
RA	Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,			
RA	Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,			
RA	Kutyavyn T., Levy R., Li M.-J., McClelland E., Palmieri A.,			
RA	Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,			
RA	Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,			
RA	Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Doian M.,			
RA	Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,			
RA	Nester E.W.;			
RT	"The genome of the natural genetic engineer Agrobacterium tumefaciens			
RT	C58.";			
RL	Science 294:2317-2323(2001).			
RN	[2]			
RN	SEQUENCE FROM N.A.			
RX	MEDLINE=21608551; PubMed=11743194;			
RA	Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,			
RA	Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,			
RA	Houmlet K., Gordon J., Vaudin M., Doughty D., Iartchouk O., Epp A., Liu F.,			
RA	Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,			
RA	Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,			
RA	Cielo C., Slater S.;			
RT	"Genome sequence of the plant pathogen and biotechnology agent			
RT	Agrobacterium tumefaciens C58.";			
RL	Science 294:2323-2328(2001).			
DR	EMBL; AE008982; AAL41112.1; -.			
DR	EMBL; AE007950; AAK85907.1; -.			
KW	Initiation factor; Complete proteome.			
SQ	SEQUENCE 913 AA; 98821 MW; 2BFICE264C484D8 CRC64;			
Query Match 11.1%; Score 146; DB 16; Length 913;				
Best Local Similarity 25.2%; Pred. No. 0.22;				
Matches 65; Conservative 34; Mismatches 79; Indels 80; Gaps				
QY	29 PPALCARP-----VKEPRGL-----SAASPPLAETGAPRRFRSVPGRG-----AAGAV 72			
Db	72 PAPVAARQAPQPRHOGGQQQRPGSQSQSGSAPQOQADRPGRVNLHDSAGEM 131			
QY	73 QELARALAHLEA-----ERQERARAEAEEDQARVLQAQLLRVYGAPRNSDPALGLD 126			
Db	132 EARRR---LMEAQADVWEAKQRAEDARRKVEEQRIAAE--KMEANRAAEAAAK 186			
QY	127 DDPDPAQAUALRLRLDPALAAQLVPAPVPAALRPRPPVDDGPGAPDAEEAGDE 186			
Db	187 VAASQPAAEV-----RAEPASERPAAAAAPRTDAPQ-----SAAAAPRSPA--- 231			
QY	187 TPDVDPPELLRYLLGRILAGSADSGVAPRRLRAADHD-----VGSLEPPEGVLGAL 239			
Db	232 TPD-----AAAPRRTGGDDEDRGAVRRGSSLPARG----- 264			
QY	240 LRVKRLETPAPQVPAERL 257			
Db	265 ----KVYAPAPAKPAERL 278			
RESULT 7				
Q9NY42	Q9NY42	PRELIMINARY:	PRT:	388 AA.

Query Match 10.28; Score 134; DB 16; Length 553;
Best Local Similarity 25.08; Pred. No. 0.72;
Matches 70; Conservative 24; Mismatches 108; Indels 78; Gaps 11;

QY 24 GLFRPPALCALCPVKEPRG-----LSAASP-----LAET 53
DB 32 GTLAPEAPVQIRPVRRAGQVPTTSIALSGVEPPAPAEVQVRYRDPDSRTTAAEEP 91
QY 54 GAPRRFRSVPGRGAGAVQELARALAHLEAEQERAR-----AEQAEDQOAR 104
DB 92 VAPTRTSKARARPTPAETVPSAPVQAPVAEAEPEKPRPTKAAAPVPAASPAETEAA 151
QY 105 VLAQLLRWGAPRNSDPALGDD---DPDAPAAQALARALLRDLPAALAAQ-----LVPAP 158
DB 152 PAS-----APRSGWLSALDLWKEPEAPAPRRAAPARQYQERSTATQTQVIOAAP 204
QY 159 VPAALRPRPEVYDDGAPGDAEE---AGDETDPDPELLRYLLGRILAGSADSEGVAAP 215
DB 205 KPAPA--PQP---DSDSLFRTQELASDLPLDPLVELLERLWEQQAQEQEEPAPAP 259
QY 216 RRLRRAADHDVGSSELPPEGVGLGALLRYKRLTTPAQVPPAR 255
DB 260 ----RAPVRPVAALSP-----APSPAAPPR 281

RESULT 10
Q9Z614 ID Q9Z614 PRELIMINARY; PRT; 612 AA.
AC Q9Z614;
DT 01-MAY-1999 (T-EMBLrel. 10, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Dihydrolipoamide acetyltransferase.
GN PDHB.
OS Streptomyces seoulensis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=73044;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99078078; PubMed=9858775;
RA Youn H., Kwak J., Youn H.D., Hah Y.C., Kang S.O.;
RT "Lipoamide dehydrogenase from streptomyces seoulensis: biochemical and genetic properties";
RL Biochim. Biophys. Acta 1388:405-418(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Youn H., Kang S.-O.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 2 LIPOYL-BINDING DOMAINS.
DR EMBL; AF047034; AAD17484.2; -;
DR HSSP; P11961; 2PDD.
DR InterPro; IPR001078; 2Oxoacid dh.
DR InterPro; IPR000089; Biotin_lipoyl.
DR InterPro; IPR004167; E3_binding.
DR InterPro; IPR003016; Lipoyl.
DR InterPro; IPR002965; P-rich_extensions.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00198; 2-oxoacid dh; 1.
DR Pfam; PF00364; biotin_lipoyl; 2.
DR Pfam; PF02817; e3_binding; 1.
DR PRINTS; PR01217; PRICHEXTENSIN.
DR PRINTS; PR01574; TUBBYPROTEIN.
DR ProDom; PD001115; 2Oxoacid dh; 1.
DR ProSITE; PS00189; LIPOYL; 2.
DR ProSITE; PS00430; TONB_DEPENDENT_REC_1; UNKNOWN_1.
KW Lipoyl; Transferase.
SQ SEQUENCE 612 AA; 61368 MW; B804109089943056 CRC64;

Query Match 10.0%; Score 132.5; DB 2; Length 612;
Best Local Similarity 27.0%; Pred. No. 0.99;
Matches 70; Conservative 21; Mismatches 117; Indels 51; Gaps 10;

QY 29 PPALCARPVKEPRGLSAAASPLAET---GAPRRFRSVPGRGAGAVQELARALAHLEA 85
DB 121 PTAAEASGGGSAECTDVLVLPALGESVTEGTVRLKSV--GDSVEADEPLLESTVDKVT 178
QY 86 ERQERARAEAE---AEDQOARVLAQLLRVWGAPRNSDPALGDDDPDAPAAQALARALLR 142
DB 179 EIPAPTSGTLLLEIVGDETAEGAKLAVIGAA--GAAPA--AAPEAPKAPAAEAPAAPAP 235
QY 143 ARLDPAALAAQLVPAPVPAALRPRP-----PVYDDGPA-----GP 178
DB 236 AAP 295
QY 179 DAEPAGDETPDPELLRYLLGRILAGSADSEGVAAPRRLRRAADHDVGSSELPPEGVLA 238
DB 296 TATQATDEGAVTBP-LVRKLAASESSVDLASVKGTVGGRIRK-----QDVIAA 342
QY 239 LLRVKRLTTPAP---QVPA 254
DB 343 AEAATAAAAPAPAAAAQAPA 361

RESULT 11
Q8RV89 ID Q8RV89 PRELIMINARY; PRT; 492 AA.
AC Q8RV89;
DT 01-JUN-2002 (T-EMBLrel. 21, Created)
DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Hypothetical 52.5 kDa protein.
GN OSJNBa0095C06.3 OR OSJNBa0091J06.20.
OS Oryza sativa (Rice), and
OS Oryza sativa (Japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530; 39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NIPPONBARE;
RA Du H., Minx P., de la Bastide M., Nascimento L., Spiegel L.,
RA Preston R., Kirchoff K., Kuit K., Baker J., Vil M.D., Zutavern T.,
RA Santos L., Cunniss D.M., Rodriguez S., Miller B., Ballija V., Shah R.,
RA Bahret A., King L., O'Shaughnessy A., Palmer L., Dedhia N.,
RA McCombie W.R.;
RT "Genomic sequence for Oryza sativa, Nipponbare strain, clone OSJNBa0095C06, from Chromosome 10, complete sequence";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA McCombie W.R., de la Bastide M., Spiegel L., Preston R., Ferraro K.,
RA Kuit K., Nascimento L., Zutavern T., Ballija V., Bell M., Baker J.,
RA Santos L., Miller B., Katzenberger F., Muller S., King L., Yang C.,
RA Dike S., O'Shaughnessy A., Palmer L., Dedhia N.;
RT "Genomic sequence for Oryza sativa, Nipponbare strain, clone OSJNBa0091J06, from chromosome 10, complete sequence";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC079935; AAM08531.1; -;
DR EMBL; AC113338; AAM08664.1; -;
KW Hypothetical protein.
SQ SEQUENCE 492 AA; 52539 MW; 23ABC4531C2587E6 CRC64;

Query Match 9.9%; Score 131; DB 10; Length 492;
Best Local Similarity 31.4%; Pred. No. 0.98;
Matches 81; Conservative 11; Mismatches 108; Indels 58; Gaps 13;

QY 18 LVLLLLGLFRPPALCARPVKEPRGL---SAAASPPLAETGAPRRFRSV-PRGEAAGAVQ 73
DB 137 LATLLPRLLCRPHRACASPPPPRLRLRPATLASPP-PELLRPRRLSVAPRSRAAPAAA 195
QY 74 ELARALAHLEAEQERARAEAEQOQARVLAQLLRVWGAPRNSDPALGDDDPDAPA 133
DB 196 RLLRSTCAALPHRRRLASSTPASRLAATLATLLPRLLR---CPHRACPA-----SPPPP- 246

```
QY 134 AQLARALLRLRDPAAALAAQLVPAPVAAALRRPVPVYDGGPAGPDAAEAGDETDFVDPE 193
DB 247 -----RLLR-----PATLASP-----PPRLRPRARLSVAPRSRAAPAAA-----R 283
QY 194 LLRYLLG-----RLAGSADSEGVAA-----PRLLRRAADHDVGSSELPPEGVLGALLRV 242
DB 284 LLRSTCAALPHRRRLASSPFAARLAATLALLPRLRCRPHRACAPSPPPPR-----LLRP 338
QY 243 KRLETPAQVPAPARRLLPP 260
DB 339 ATLASPPP-----RLLRP 351

RESULT 12
Q9NEA7 PRELIMINARY; PRT; 1290 AA.
AC Q9NEA7:
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE Hypothetical protein 136.2 kDa protein.
GN L5515.05.
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RA Masny D., Purnelle B., Goffeau A., Ivens A.C., Quail M.,
RA Rajandream M.A., Barrell B.G.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RX MEDLINE=98146435; PubMed=9477341;
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RA Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome.";
RL Genome Res. 8:135-145(1998).
DR EMBL; AL161414; CAB7679.1; -.
KW Hypothetical protein.
SQ SEQUENCE 1290 AA; 136159 MW; 6A6F8B395CE6818 CRC64;

Query Match 9.9%; Score 131; DB 5; Length 1290;
Best Local Similarity 26.6%; Pred. No. 2.6;
Matches 69; Conservative 23; Mismatches 95; Indels 72; Gaps 12;

QY 27 RPPALCARPVKEPR-GLSAASPPLAETGAP-----RRF 59
DB 95 RDPALLHTVHAQQHGVSAKGRARHSAPVGPDRSTDASDFSTVPFSLRLEVERRF 154
QY 60 RR--SVPRGEAGAVQ-ELARALAHLLAEQERARAEQAEQOARVLAQLLRVWGAP 116
DB 155 RQMKVHGLTAASQVDIMERSVOLFOER--RTRFAQALAEQDAIWA---RWV--P 207
QY 117 RNSDPALGLDDPDAPAQLARALLRLRDPAAALAAQLVPAPVAAALRRPVPVYDGGPA 176
DB 208 GCSLEK---TETETAHQAAASARVAMPASRRSDSSATPSASTAISPPPPA---SPT 261
QY 177 GPDAEAGDETPDVPPELLRYLLGRILAGSADSEGVAPRRLRRRAADHDVGSSELPPEGVL 236
DB 262 G-----FLKDL-----DTYSKPRELREAS-----GGSRSAAAAA 292
QY 237 GALLRVKRLTPAQVPAPAR 255
DB 293 PAASKATKMPKGVFLAPAR 311

RESULT 13
Q9A718 PRELIMINARY; PRT; 589 AA.
AC Q9A718;
```

```
DT 01-JUN-2001 (TremBLrel. 17, Created)
DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Hypothetical protein CC1908.
GN CC1908.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeRubeis R.T., Dodson R.J., Durkin A.S., Gilm M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005864; AAK23883.1; -.
DR TIGR; CC1908; -.
DR PRINTS; PR01656; VACCYTOTOXIN.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 589 AA; 59269 MW; A7E52050C891B98A CRC64;

Query Match 9.9%; Score 130.5; DB 16; Length 589;
Best Local Similarity 29.8%; Pred. No. 1.3;
Matches 82; Conservative 15; Mismatches 91; Indels 87; Gaps 15;

QY 45 AASPPLAETGAPRRFRSVPGEAAGAVQELARAL-AHLLEAEQERARA-----93
DB 182 AAPSASVGAAP-----PWAEGDLKALWLRLSALSAMLAKEVSSGAVAPHVGVADGG 237
QY 94 -----EAQEAEDQARVLAQLLRVVGAPRNSDPALG--LDLDDPDAPAA---QLARALLRA 143
DB 238 GADGPDPOGPMTHQARANGSSSSA-GAMASSPPAQSRLLAADAPAPATSSPIARAL---293
QY 144 RLDPAAALAAQLVPAPVAAALRRPVPVYDGGPAGDAEAGDETDFVDPEL-LRY-----197
DB 294 -----PAGAPVDAEDPPPARVGAPPSG-----SPTTHADPELAVRFGAFVA 335
QY 198 -----LLGRILAGSA-----DSEG-----VAAAPRRLRR-----AARDVGS 228
DB 336 PPKTAERSPKGVAGSLPALVQLPEPRDGESEPPATEPAEPRLMTRGYCAPVEDVRS 395
QY 229 ELPPGEGVLALLRVKRLR-TP-----APQVPARRLL 258
DB 396 KTPPPPYAGGPTAGORPEPTPSLSWAPEAMARRLL 430

RESULT 14
Q9EU24 PRELIMINARY; PRT; 350 AA.
AC Q9EU24;
DT 01-MAR-2001 (TremBLrel. 16, Created)
DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TremBLrel. 16, Last annotation update)
DE Hypothetical 37.5 kDa protein.
GN ORF31.
OS Corynebacterium equi (Rhodococcus equi).
OC Plasmid PREAT701 (p33701), and Plasmid virulence.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.
OX NCBI_TaxID=43767;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC33701; PLASMID=PREAT701 (p33701);
RA Takai S., Sekizaki T., Kakuda T., Nakamura M., Suzuki K., Ogino N.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
```



```

RP SEQUENCE FROM N.A.
RC STRAIN=ATCC33701, AND 103; PLASMID=PREAT701 (P33701), AND VIRULENCE;
RX Pubmed=11083803;
RA Takai S., Hines S.A., Sekizaki T., Nicholson V.M., Alperin D.A.,
RA Osaki M., Takamatsu D., Nakamura M., Suzuki K., Ogino N., Kakuda T.,
RA Dan H., Prescott J.F.;
RA "DNA sequence and comparison of virulence plasmids from Rhodococcus
RT equi ATCC 33701 and 103.";
RL Infect. Immun. 68:6840-6847(2000).
RL EMBL; AP001204; BAB16640.1; -.
DR EMBL; AF116907; AAG21734.1; -.
SW Hypothetical protein; Plasmid.
SQ SEQUENCE 350 AA; 37537 MW; 56A737F3318B62C3 CRC64;

Query Match 9.9%; Score 130; DB 2; Length 350;
Best Local Similarity 28.4%; Pred. No. 0.8;
Matches 79; Conservative 20; Mismatches 89; Indels 90; Gaps

QY 27 RPPALCARPVKEPRGLSASPLAETGAP-----AHLEAEQERARAEAEAEED--- 100
|| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 12 RPORARAIDP--GPDGAAAPQPDWAKRGRPAHEQAHQDVVDVAGFLSKVDRAAGRSVP 69
65 R-----GEAAGA-----VQELARAL-----AHLEAEQERARAEAEAEED--- 100
70 KPQPMREALGAFHRSRDPDQTRVQVLAATAARHLEGAQSAQAQRWERLARENADRLDALR 129
101 -----QOARYLAQLLRVWGAPRNSDPALGLDDPDPA--PAAQLARALLRAR-LD 146
130 ROOHREQVAYQAARMSPEIQOQDQRASET---ESDKRAFDAANAALGVVAARGLP 186
147 P-----AALAAQL--VPAPVPAAL--RPRPVYVDGSPAGPDAAE--AGDETDPDPELLR 196
| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 187 PFDRLASAEHLDDVPAPEPTFEVDEPTADVQSDPSASPDPEQGPAADEAVDV----- 240
197 YLGLRILAGSADSEGVAAAPRRLRRADHDVGSFLPPEG 234
|| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 241 -----SADLEPVATPEAGGASAPSRGLPELERAG 269

RESULT 15
Q8VCG2 PRELIMINARY; PRT; 544 AA.
AC Q8VCG2;
DT 01-WAR-2002 (TReMBLrel. 20, Created)
DT 01-WAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-WAR-2002 (TReMBLrel. 20, Last annotation update)
DE Hypothetical 59.8 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SPALIVARY GLAND;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
RL EMBL; BC019977; AAH19977.1; -.
DR Hypothetical protein.
KW Hypothetical protein.
SQ SEQUENCE 544 AA; 59840 MW; 075828F5B370111A CRC64;

Query Match 9.8%; Score 129.5; DB 11; Length 544;
Best Local Similarity 25.9%; Pred. No. 1.3;
Matches 74; Conservative 26; Mismatches 117; Indels 69; Gaps

QY 2 AGSPLLWGRPAGVGLLVLLLLGLFRPPALCARPVKEPRGLSASPLAETGAPRFR 61
|| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 39 AGASLAPGPHD-----RTHPS-AAVPVC-PR--SASASPLTPCSAPRSAGR 81
62 SVPRGE-----AAGAVQELARALAHLEAEQERARAEAEAEQOARVLAQ---LLR 111
|| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 82 CTPSGERPERKPCAGGSPALARRULEATPVQKEKKEKEKEKSA--LARENKLR 139
112 VWGAPRNSDPALGLDD-----PDAPAAQLARALLRARLDPAALAAQLVPAVPVAAA 163

```


GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 12, 2003, 16:26:07 ; Search time 2721 Seconds
(without alignments)
10471.021 Million cell updates/sec

Title: US-09-803-589-5

Perfect score: 979

Sequence: 1 gaattcgacagaggccagc.....gaggatccctaccctggc 979

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba:*
- 2: gb_hgt:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pi:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_srs:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vi:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_lm:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pi:*
- 26: em_ro:*
- 27: em_srs:*
- 28: em_un:*
- 29: em_vi:*
- 30: em_hgt_hum:*
- 31: em_hgt_inv:*
- 32: em_hgt_other:*
- 33: em_hgt_mus:*
- 34: em_hgt_pln:*
- 35: em_hgt_rod:*
- 36: em_hgt_nam:*
- 37: em_hgt_vrt:*
- 38: em_sy:*
- 39: em_hgo_hum:*
- 40: em_hgo_mus:*
- 41: em_hgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	959.8	98.0	998	6	AX083428	Sequence
2	941.6	96.2	993	9	BC002851	AX083428 Homo sapi
3	940	96.0	969	9	AF181562	AX083428 Homo sapi
4	776.8	79.3	780	6	AX083418	Sequence
5	599	61.2	1027	10	AF181560	AF181560 Mus muscu
6	583	59.6	945	10	AF181561	AF181561 Rattus no
7	580.2	59.3	975	10	BC012263	BC012263 Mus muscu
8	502.2	51.3	1025	10	AF293356	AF293356 Mus muscu
9	474.2	48.4	113853	9	AF196971	AF196971 Homo sapi
10	343	35.0	62396	2	AC098339	AC098339 Rattus no
11	343	35.0	77876	2	AC098473	AC098473 Rattus no
12	343	35.0	111948	2	AC094563	AC094563 Rattus no
13	332.6	34.0	151694	10	AL670169	AL670169 Mouse DNA
14	308.6	31.5	177613	2	AC123014	AC123014 Rattus no
15	276.6	28.3	301	6	AX336215	AX336215 Sequence
16	276.6	28.3	301	6	AX409054	AX409054 Sequence
17	143.6	14.7	77876	2	AC098473	AC098473 Rattus no
18	122.4	12.5	177613	2	AC123014	AC123014 Rattus no
19	99	10.1	210215	2	AC097157	AC097157 Rattus no
20	98.2	10.0	206773	2	AC126296	AC126296 Rattus no
21	93.2	9.5	74721	2	AC096161	AC096161 Rattus no
22	92.2	9.4	991	11	PM12H12B	AL684455 Penicilli
23	92	9.4	136551	2	AC048354	AC048354 Homo sapi
24	91.4	9.3	174410	2	AC128158	AC128158 Rattus no
25	91	9.3	184402	2	AC127041	AC127041 Rattus no
26	90.6	9.3	74721	2	AC096161	AC096161 Rattus no
27	90.2	9.2	110958	2	AC118521	AC118521 Rattus no
28	90	9.2	1393	11	PM11H12G	AL684456 Penicilli
29	89.8	9.2	151730	2	AC125640	AC125640 Rattus no
30	88.6	9.1	151996	2	AC126885	AC126885 Rattus no
31	88.6	9.1	167527	2	AC121745	AC121745 Rattus no
32	88.2	9.0	253464	2	AC11916	AC11916 Rattus no
33	88	9.0	93419	2	AC121747	AC121747 Rattus no
34	87.8	9.0	135733	2	AC128114	AC128114 Rattus no
35	87.8	9.0	158897	2	AC121439	AC121439 Rattus no
36	87.6	8.9	10444	14	PW1422133	AJ422133 Pseudorab
37	87.4	8.9	85268	2	AC022648	AC022648 Homo sapi
38	87.4	8.9	158897	2	AC121439	AC121439 Rattus no
39	86.8	8.9	1094	11	PW7G11B	AL685196 Penicilli
40	86.6	8.8	141905	2	AC110101	AC110101 Rattus no
41	86.6	8.8	147622	2	AC110437	AC110437 Rattus no
42	86.4	8.8	61617	2	AC120284	AC120284 Rattus no
43	85	8.7	956	11	PM2D12B	AL684743 Penicilli
44	85	8.7	35487	2	AC110453	AC110453 Rattus no
45	85	8.7	52683	2	AC131481	AC131481 Rattus no

ALIGNMENTS

RESULT 1	AX083428	Sequence 120 from Patent WO0112660.	998 bp	DNA	linear	PAT 28-FEB-2001
LOCUS	AX083428	Sequence 120 from Patent WO0112660.				
DEFINITION	AX083428					
ACCESSION	AX083428					
VERSION	AX083428.1	GI:13185274				
KEYWORDS	human.					
SOURCE	human.					
ORGANISM	Homo sapiens					
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;					
	Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.					
REFERENCE	1 (bases 1 to 998)					
AUTHORS	Kato,S. and Kimura,T.					
TITLE	Human proteins having hydrophobic domains and dnas encoding these proteins					

JOURNAL Patent: WO 0112660-A 120 22-FEB-2001;
SAGAMI CHEMICAL RESEARCH CENTER (JP) ; Protegene Inc. (JP)
FEATURES
source
1. .998
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
50. .832
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC33311.1"
/db_xref="GI:13185275"
/translation="MAGSPLLMGPRAGVGLVLLGLRPPPPALCARPVKEPRGLS
AASPLAETGAPRRFRSVPRGEAGVQELARALAHLEERGERARAEQEDQ
ARVLAQLRWVAGPRNSDPALGLDDPDPAQAQLRALRLDPAALAAQLVPAVP
AALRRPPVDDGPDAGDEAGDETDPVDPELLRYLLGRILLAGSADSEGVAAAPRL
RRAADHDVGSSELPEGVIGALLRVKLETPAPQVPAARRLLPP"
BASE COUNT 120 a 394 c 347 g 137 t
ORIGIN

Query Match 98.04; Score 959.8; DB 6; Length 998;
Best Local Similarity 99.34; Pred. No. 2.9e-104;
Matches 958; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 15 GCCAGCCAGTCCGCGSGYMCGRBCCCGCTCGCTGGGCGCATGGCGGGTCCGCCGT 74
|||||
Db 7 GCCAGCCAGTCCGCGCGGAGCCCGCTCGCTGGGCGCATGGCGGGTCCGCCGT 66
|||||
Qy 75 GCTCTGGGGCGCGCGGGCGGCTCGGCTTTTGTGCTGCTGCTGCTGCGGCTGTT 134
|||||
Db 67 GCTCTGGGGCGCGGGCGGCGGCTTTTGTGCTGCTGCTGCTGCTGCGGCTGTT 126
|||||
Qy 135 TCGGCGCGCGCGCGGCTGCTGCGCGCGCGGTAAGAGAGCGCGCGGCTTAAGCGCAGC 194
|||||
Db 127 TCGGCGCGCGCGCGGCTGCTGCGCGCGCGGTAAGAGAGCGCGCGGCTTAAGCGCAGC 186
|||||
Qy 195 GTCTCGCGCTTGGCTGAGCTGGCGCTTCGCGCTTCGCGGCTCAGTGCCCGGAGG 254
|||||
Db 187 GTCTCGCGCTTGGCTGAGCTGGCGCTTCGCGCTTCGCGGCTCAGTGCCCGGAGG 246
|||||
Qy 255 TGAGCGCGCGGGCGGTCGAGGAGCTGGCGCGGGCGCTGGCGCATCTGCTGGAGCGCGA 314
|||||
Db 247 TGAGCGCGGGCGGTCGAGGAGCTGGCGCGGGCGCTGGCGCATCTGCTGGAGCGCGA 306
|||||
Qy 315 ACCTCAGAGCGGGCGGGCGGCGGAGCGCAGAGCTGAGATCAGCAGGCGCGGCTCCT 374
|||||
Db 307 ACCTCAGAGCGGGCGGGCGGCGGAGCGCAGAGGCTGAGATCAGCAGGCGCGGCTCCT 366
|||||
Qy 375 GCGCAGCTGCTGCGCTGCTGGGCGCGCCCGCAACTCTGATCGGCTCTGGGCTTGA 434
|||||
Db 367 GCGCAGCTGCTGCGCTGCTGGGCGCGCCCGCAACTCTGATCGGCTCTGGGCTTGA 426
|||||
Qy 435 CGACGACCCCGACGCGCTGCGAGGCGAGCTGCTGCGCTGCTGCTGCGCGCGCGCTTGA 494
|||||
Db 427 CGACGACCCCGACGCGCTGCGAGGCGAGCTGCTGCGCTGCTGCTGCGCGCGCGCTTGA 486
|||||
Qy 495 CCCTGCGCGCTTACAGCCAGCTTTGCTCCCGCGCGCTTCCCGCGGGCGCTCCGACC 554
|||||
Db 487 CCCTGCGCGCTTACAGCCAGCTTTGCTCCCGCGCGCTTCCCGCGGGCGCTCCGACC 546
|||||
Qy 555 CCGCGCCCGCTTACAGCAGCGCGCGCGCGCGCGGATCTGAGGAGGCGAGCGAGCA 614
|||||
Db 547 CCGCGCCCGCTTACAGCAGCGCGCGCGCGCGGATCTGAGGAGGCGAGCGAGCA 606
|||||
Qy 615 GACACCGACGTCGACCCCGAGCTTTGAGGTACTTGTGGGAGCGGATTTCTGGGGGAG 674
|||||
Db 607 GACACCGACGTCGACCCCGAGCTTTGAGGTACTTGTGGGAGCGGATTTCTGGGGAG 666
|||||
Qy 675 CGCGAGCTCCGAGGGGTGGCAGCGCGCGCGCGCTCCGCGCGCGCGACAGATGT 734
|||||
Db 667 CGCGAGCTCCGAGGGGTGGCAGCGCGCGCGCGCTCCGCGCGCGCGACAGATGT 726
|||||
Qy 735 GGGCTCTAGCTGCCCTTGGGCGGCTGCTGGGCGGCTGCTGCTGTAACGCTAGA 794
|||||

Db 727 GGGCTCTGAGCTGCCCCCTGAGGGCGTGTGGGGGGCGTGTGTGTGAACGCTAGA 786
Qy 795 GACCCCGCGCGCCAGGTCCTGTGCACGCGCGCTCTTTGGCACACCTGAGCACTGCCCGGATC 854
|||||
Db 787 GACCCCGCGCGCCAGGTCCTGTGCACGCGCGCTCTTTGGCACACCTGAGCACTGCCCGGATC 846
|||||
Qy 855 CCGTGCACCTTGGGACCCAGAGTGTCCCGCGCATCCCGCCACACAGGACTGTCTCCCGCC 914
|||||
Db 847 CCGTGCACCTTGGGACCCAGAGTGTCCCGCGCATCCCGCCACACAGGACTGTCTCCCGCC 906
|||||
Qy 915 AGCAGTCCAGCAACTTACCCCGCGCGCGCGCTCTACCCGAGGATCCCTTACCC 974
Db 907 AGCAGTCCAGCAACTTACCCCGCGCGCGCGCTCTACCCGAGGATCCCTTACCC 966
Qy 975 CTGGC 979
|||||
Db 967 CTGGC 971

BC002851 993 bp mRNA linear PRI 12-JUL-2001
Homo sapiens, granin-like neuroendocrine peptide precursor, clone
MGC:3466 IMAGE:3660001, mRNA, complete cds.
ACCESSION BC002851
VERSION BC002851.1 GI:12804000
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (05-FEB-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK
COMMENT NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland:
Web site: <http://www.nisc.nih.gov/>
Contact: nisc_mgc@nigr.nih.gov
Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakesley, R.W., Bouffard, G., Brinkley, C., Brooks, S.,
Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,
Lim, M., Maduro, Q.L., Mastello, C., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J.,
Thompson, E.E., Touchman, J.W., Tsugeon, C., Vogt, J.L., Walker, M.A.,
Zhang, L.-H. and Green, E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 12 Row: n Column: 24
This clone was selected for full length sequencing because it
passed the following selection criteria: GenomeScan gene
prediction.
FEATURES
LOCATION/Qualifiers
1. .993
/organism="Homo sapiens"
/db_xref="LocusID:27344"
/db_xref="taxon:9606"
/clone="MGC:3466 IMAGE:3660001"
/tissue_type="Uterus, endometrium adenocarcinoma"
/clone_lib="NIH_MGC_44"
/lab_host="DH10B-R"
/note="Vector: pOTB7"

```
CDS      30..812
/codon_start=1
/product="granin-like neuroendocrine peptide precursor"
/db_xref="AAH02851.1"
/db_xref="GI:12804001"
/translation="MAGSPLLMGPRAGGVLLVLLGLFRPPALCARVKEPRGLS
AASPLAETGAPRRFRSVPGEAGAVOELARALHLEAEQERARAEADQO
ARVLAQLRWGAPRNSDPALGLDDPDPAQAOLARALLARLDPAALAAQLVPAPVP
AALRRPVPYDDGPADEAGDETDPDPELLRLLGLRILAGSADSEGVAAAPRL
RRAADHDVGSSELPPGVLGALLRKRLTAPQVPAARRLLP"
BASE COUNT      136 a      381 c      343 g      133 t
ORIGIN
Query Match      96.2%; Score 941.6; DB 9; Length 993;
Best Local Similarity 99.7%; Pred. No. 4e-102;
Matches 941; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 36 RRGCCCGCTCGCTGGGCGACATGGCGGGTGCCTGCTGCGGGCGCGGGCGG 95
Db 8 GGGCCCGCTCGCTGGGCGACATGGCGGGTGCCTGCTGCGGGCGCGGGCGG 67
QY 96 GGGCGTCGGCTTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 155
Db 68 GGGCGTCGGCTTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 127
QY 156 CGCGCGCGCGTAAAGAGCGCGCGCGCTAAGCGCAGCGCTCTCGCGCTTGG 215
Db 128 CGCGCGCGCGTAAAGAGCGCGCGCGCTAAGCGCAGCGCTCTCGCGCTTGG 187
QY 216 TGGCGCTCTCGCGCTTCCGGCGGTGAGTGCCTCGCGGTGAGCGCGGGGCG 275
Db 188 TGGCGCTCTCGCGCTTCCGGCGGTGAGTGCCTCGCGGTGAGCGCGGGGCG 247
QY 276 GGAGCTGGCGCGCGCTGGCGATCTGCTGGAGCGCGAAGCTGAGAGCGGGCG 335
Db 248 GGAGCTGGCGCGCGCTGGCGATCTGCTGGAGCGCGAAGCTGAGAGCGGGCG 307
QY 336 CGAGGCGCAGAGCTGAGATGAGCAGGCGCGCGTCTGCGCGAGCTGCTGCG 395
Db 308 CGAGGCGCAGAGCTGAGATGAGCAGGCGCGCGTCTGCGCGAGCTGCTGCG 367
QY 396 GGGCGCGCGCGCGCTGATCGCGCTGCGGTGCGAGCGACCGCGCGCGCTGC 455
Db 368 GGGCGCGCGCGCGCTGATCGCGCTGCGGTGCGAGCGACCGCGCGCGCTGC 427
QY 456 AGCGAGCTCGCTGCGCTGCTGCTCGCGCGCGCTGAGCTGCTGCGCGCTGAG 515
Db 428 AGCGAGCTCGCTGCGCTGCTGCTCGCGCGCGCTGAGCTGCTGCGCGCTGAG 487
QY 516 GCTTGTCCCGCGCGCTGCGCGCGCGCTGCGAGCGACCGCGCGCGCTGAG 575
Db 488 GCTTGTCCCGCGCGCTGCGCGCGCGCTGCGAGCGACCGCGCGCGCTGAG 547
QY 576 CGGCGCGCGCGCGCGCTGAGGAGCGAGCGAGCAGCAGCAGCGTGGACCCCG 635
Db 548 CGGCGCGCGCGCGCGCTGAGGAGCGAGCGAGCAGCAGCAGCGTGGACCCCG 607
QY 636 GCTGTTGAGTACTTGTGGAGCGATTTGCGGGAGCGCGAGCTCCGAGGGGGT 695
Db 608 GCTGTTGAGTACTTGTGGAGCGATTTGCGGGAGCGCGAGCTCCGAGGGGGT 667
QY 696 AGCGCGCGCGCGCTGCGCGCGCGCGAGCAGATGTGGGTCTGAGTGGCCCTG 755
Db 668 AGCGCGCGCGCGCTGCGCGCGCGCGAGCAGATGTGGGTCTGAGTGGCCCTG 727
QY 756 GGGGTGCTGGGGCGCTGCTGGGTGTGAACGCTAGACCGCGCGCGCGAGGTGC 815
Db 728 GGGGTGCTGGGGCGCTGCTGGGTGTGAACGCTAGACCGCGCGCGCGAGGTGC 787
QY 816 TGCACGCGCGCTTGTGACCGCTGAGCAGCTGCGCGGATCCCGTGCACCTGG 875
Db 788 TGCACGCGCGCTTGTGACCGCTGAGCAGCTGCGCGGATCCCGTGCACCTGG 847
```

```
QY 876 AGTGCCCGCGCCATCCGCCACAGGACTGCTCCCGCCAGCAGCTCCAGAGCAACTTAC 935
Db 848 AGTGCCCGCGCCATCCGCCACAGGACTGCTCCCGCCAGCAGCTCCAGAGCAACTTAC 907
QY 936 CCGCGCCAGCAGCGCTCTCACCAGGATCCCTACCCCTGGC 979
Db 908 CCGCGCCAGCAGCGCTCTCACCAGGATCCCTACCCCTGGC 951
RESULT 3
AF181562 969 bp mRNA linear PRI 27-JAN-2000
LOCUS Homo sapiens proSAAS mRNA, complete cds.
DEFINITION AF181562
ACCESSION AF181562
VERSION AF181562.1 GI:6653212
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 969)
AUTHORS Fricker,L., McKinzie,A.A., Sun,J., Curran,E., Qian,Y., Yan,L.,
Patterson,S.D., Courchesne,P.L., Richards,B., Levin,N., Mzhavia,N.,
Devi,L.A. and Douglass,J.
TITLE Identification and characterization of proSAAS, a granin-like
neuroendocrine peptide precursor that inhibits prohormone
processing
JOURNAL J. Neurosci. 20 (2), 639-648 (2000)
MEDLINE 20098938
PUBMED 10632593
REFERENCE 2 (bases 1 to 969)
AUTHORS Fricker,L., McKinzie,A.A., Sun,J., Curran,E., Qian,Y., Yan,L.,
Patterson,S.D., Courchesne,P.L., Richards,B., Levin,N., Mzhavia,N.,
Devi,L.A. and Douglass,J.
TITLE Direct Submission
JOURNAL Submitted (30-AUG-1999) Department of Molecular Pharmacology,
Albert Einstein College of Medicine, 1300 Morris Park Avenue,
Bronx, NY 10461, USA
FEATURES
source Location/Qualifiers
1..969
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="FLHSAAS"
1..26
/feature="proSAAS"
27..809
/function="Inhibits prohormone processing"
/notes="granin-like neuroendocrine peptide precursor"
/codon_start=1
/product="proSAAS"
/protein_id="AAF22643.1"
/db_xref="GI:6653213"
/translation="MAGSPLLMGPRAGGVLLVLLGLFRPPALCARVKEPRGLS
AASPLAETGAPRRFRSVPGEAGAVOELARALHLEAEQERARAEADQO
ARVLAQLRWGAPRNSDPALGLDDPDPAQAOLARALLARLDPAALAAQLVPAPVP
AALRRPVPYDDGPADEAGDETDPDPELLRLLGLRILAGSADSEGVAAAPRL
RRAADHDVGSSELPPGVLGALLRKRLTAPQVPAARRLLP"
sig_peptide 27..125
3'UTR 810..969
BASE COUNT 117 a 379 c 338 g 135 t
ORIGIN
Query Match 96.0%; Score 940; DB 9; Length 969;
Best Local Similarity 99.2%; Pred. No. 6.2e-102;
Matches 940; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
QY 32 YMCRRRCGCGCTCGCTGGGCGAGCATGGCGGTGCTGCTGCTGGGGCGCGGG 91
Db 1 TCCGGAGCAGGCTCGCTGGGCGAGCATGGCGGTGCTGCTGCTGGGGCGCGGG 60
QY 92 CCGGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 151
Db 61 CCGGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120
```

QY	152	TCCTGGCGCGCGGTAAGGAGCCCGCGGCTAAGCCAGCGTCTCGGCCCTTGGCTG	211
Db	121	TCCTGGCGCGCGGTAAGGAAACCGCGGCTAAGCCAGCGTCTCGGCCCTTGGCTG	180
QY	212	AGACTGGCGGCTCTCGCGCTTCCGCGGTAGTCGCCGAGGTGAGCGGCGGGGGGG	271
Db	181	AGACTGGCGGCTCTCGCGCTTCCGCGGTAGTCGCCGAGGTGAGCGGCGGGGGGG	240
QY	272	TCAGAGAGTGGCGGCGGCTGGGCGATCTGCTGGAGGCCGAAGCTCAGGAGCGGGCG	331
Db	241	TCAGAGAGTGGCGGCGGCTGGGCGATCTGCTGGAGGCCGAAGCTCAGGAGCGGGCG	300
QY	332	GGGCGAGCGCAGAGAGGTGAGGATCAGCAGCGCGGTCTCTGGCGAGCTGCTGGCG	391
Db	301	GGGCGAGCGCAGAGAGGTGAGGATCAGCAGCGCGGTCTCTGGCGAGCTGCTGGCG	360
QY	392	TCCTGGGCGCGCGCGCGAACTCTGATCGGCTTGGGCTTGGACACGACACCCGACGGC	451
Db	361	TCCTGGGCGCGCGCGCGAACTCTGATCGGCTTGGGCTTGGACACGACACCCGACGGC	420
QY	452	CTGACGCGAGCTCGCTCGCGCTCTGCTCCGCGCGCGCTTGACCTTGCCTAGCAG	511
Db	421	CTGACGCGAGCTCGCTCGCGCTCTGCTCCGCGCGCGCTTGACCTTGCCTAGCAG	480
QY	512	CCGAGCTTGTCCCGCGCGCTCCCGCGCGCGCGCTCGGACCCCGCGCGCTTACG	571
Db	481	CCGAGCTTGTCCCGCGCGCGCTCCCGCGCGCGCGCTCGGACCCCGCGCGCTTACG	540
QY	572	ACGACGGCGCGCGCGCGGATCTGAGGAGGAGGAGGACGACACCCGACGTGGAC	631
Db	541	ACGACGGCGCGCGCGCGGATCTGAGGAGGAGGAGGACGACACCCGACGTGGAC	600
QY	632	CCGAGCTGTGAGGTACTTGTGGACGATTTCTCGGGAAGCGGACCTCCGAGGGG	691
Db	601	CCGAGCTGTGAGGTACTTGTGGACGATTTCTCGGGAAGCGGACCTCCGAGGGG	660
QY	692	TGGCAGCGCGCGCGCTCCCGCGCGCGCGCGGACGATGTGGGCTCTGAGCTGCC	751
Db	661	TGGCAGCGCGCGCGCTCCCGCGCGCGCGGACGATGTGGGCTCTGAGCTGCC	720
QY	752	CTGAGGCGGCTGTGGGCGGCTGCTGCGGTGAACGCTAGAGACCCGCGCGCGCAGG	811
Db	721	CTGAGGCGGCTGTGGGCGGCTGCTGCGGTGAACGCTAGAGACCCGCGCGCGCAGG	780
QY	812	TGCTGACGCGCGCTCTTTCACCTTGCACCTGAGCACTGCCCGGATCCCGTGGGAC	871
Db	781	TGCTGACGCGCGCTCTTTCACCTTGCACCTGAGCACTGCCCGGATCCCGTGGGAC	840
QY	872	CAGAGTGCCTCCCGCATCCCGCACGAGGACTGCTCCCGCGCAGCAGCTCCAGGAA	931
Db	841	CAGAGTGCCTCCCGCATCCCGCACGAGGACTGCTCCCGCGCAGCAGCTCCAGGAA	900
QY	932	TTACCCCGCGCAGCGCGCTCTACCGGAGGATCCCTACCCCTTGGC	979
Db	901	TTACCCCGCGCAGCGCGCTCTACCGGAGGATCCCTACCCCTTGGC	948
RESULT 4	AX083418	Sequence 110 from Patent WO0112660.	
LOCUS	AX083418	780 bp DNA linear	PAT 28-FEB-2001
DEFINITION	AX083418	Sequence 110 from Patent WO0112660.	
ACCESSION	AX083418	Sequence 110 from Patent WO0112660.	
VERSION	AX083418.1	GI:13185255	
KEYWORDS	human.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
TITLE	1 (bases 1 to 780)		
JOURNAL	Kato, S. and Kimura, T.		
	Human proteins having hydrophobic domains and dnas encoding these		
	proteins		
	Patent: WO 0112660-A 110 22-FEB-2001;		

SAGAMI CHEMICAL RESEARCH CENTER (JP) ; Protegene Inc. (JP)	
FEATURES	Location/Qualifiers
source	1..780
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
BASE COUNT	79 a 297 c 294 g 110 t
ORIGIN	
Query Match	79.3%; Score 776.8; DB 6; Length 780;
Best Local Similarity	99.7%; Pred. No. 1e-82;
Matches	778; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY	58 ATGCGGGGTGCGCGCTCTCTGGGGGCGCGGGGCGGCGCTCTGGGCTTTTGGTGTG 117
Db	1 ATGCGGGGTGCGCGCTCTCTGGGGGCGCGGGGCGGCGCTCTGGGCTTTTGGTGTG 60
QY	118 CTGCTGCTGGGCTGTTTGGCGCGCGCGCGCGCTCTGCGCGCGCGCGGTAAGAGAGCCC 177
Db	61 CTGCTGCTGGGCTGTTTGGCGCGCGCGCGCGCTCTGCGCGCGCGCGGTAAGAGAGCCC 120
QY	178 CGGGGCTTAAGCGCAGCGTCTCGGCCCTTGCGCTGAGACTGGCGCTCTCGCGCGCTCCGG 237
Db	121 CGGGGCTTAAGCGCAGCGTCTCGGCCCTTGCGCTGAGACTGGCGCTCTCGCGCGCTCCGG 180
QY	238 CGGTCACTGCCCGAGGTGAGGGGCGGGGCGGTGTCAGAGACTGGCGGGCGCTGGCG 297
Db	181 CGGTCACTGCCCGAGGTGAGGGGCGGGGCGGTGTCAGAGACTGGCGGGCGCTGGCG 240
QY	298 CATCTGCTGAGGCGCGAAGCTCAGAGAGCGGGCGGGCGGAGGCGCAGAGGCTGAGGAT 357
Db	241 CATCTGCTGAGGCGCGAAGCTCAGAGAGCGGGCGGGCGGAGGCGCAGAGGCTGAGGAT 300
QY	358 CAGCAGCGCGCTCTCTGGCGGCGAGCTGCTGGGCTCTGGGGCGGGCGGCAACTCTGAT 417
Db	301 CAGCAGCGCGCTCTCTGGCGGCGAGCTGCTGGGCTCTGGGGCGGGCGGCAACTCTGAT 360
QY	418 CCGGCTCTGGGCTTGGACGACGACCCGACGCGCTTCAGCAGAGCTCGCTCGCGCTCTG 477
Db	361 CCGGCTCTGGGCTTGGACGACGACCCGACGCGCTTCAGCAGAGCTCGCTCGCGCTCTG 420
QY	478 CTCGGCGCGCGCTTGACCTTGCCTTGCCTTGCCTTGCCTTGCCTTGCCTTGCCTTGCCT 537
Db	421 CTCGGCGCGCGCTTGACCTTGCCTTGCCTTGCCTTGCCTTGCCTTGCCTTGCCTTGCCT 480
QY	538 GCCGCGCGCTCCGACCCCGCGCTTACGACGACGCGCGCGCGCGCGCGGATGCT 597
Db	481 GCCGCGCGCTCCGACCCCGCGCTTACGACGACGCGCGCGCGCGCGCGGATGCT 540
QY	598 GAGGAGCAGCGCAGAGACACCCGACGCTGGACCCCGAGCTGTTGAGTACTTGTGGGA 657
Db	541 GAGGAGCAGCGCAGAGACACCCGACGCTGGACCCCGAGCTGTTGAGTACTTGTGGGA 600
QY	658 CGGATTTCTGGGGAAGCGGAGCTCCGAGGGGTGGCAGCCCGCGCGCTCCGCGCT 717
Db	601 CGGATTTCTGGGGAAGCGGAGCTCCGAGGGGTGGCAGCCCGCGCGCTCCGCGCT 660
QY	718 GCCGCGCGACCATGTGGGCTCTGAGCTGCCCCCTTGAAGGCGTGTGGGGGCGCTGCTG 777
Db	661 GCCGCGCGACCATGTGGGCTCTGAGCTGCCCCCTTGAAGGCGTGTGGGGGCGCTGCTG 720
QY	778 CGTGTGAACCGCTTAGAGACCCCGCGCGCGCGCGCTGCTGACGCGCGCTCTTGGCACACC 837
Db	721 CGTGTGAACCGCTTAGAGACCCCGCGCGCGCGCGCTGCTGACGCGCGCTCTTGGCACACC 780
RESULT 5	
AF181560	
LOCUS	AF181560 1027 bp mRNA linear ROD 27-JAN-2000
DEFINITION	Mus musculus proSAAs gene, complete cds.
ACCESSION	AF181560
VERSION	AF181560.1 GI:6653208
KEYWORDS	
SOURCE	Mus musculus.

FEATURES

```
source      1. .945
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="FLRSAAS"
1. .21
22. .804
/function="inhibits prohormone processing"
/notes="granin-like neuroendocrine peptide precursor"
/codon_start=1
/product="proSAAS"
/protein_id="AAE22642.1"
/db_xref="GI:6653211"
/translaton="MAGSPLLCPRAGGVGLVLLGLLRLPPTLSARVVKPRPSLS
AASPLAETPLRLRAVPRGEAGAVOELARALAHLEAERARARAOEAEDQO
ARYLAQLRWGSPRASDPPLADDDPDAQAOLARALLRLRDPALAAQLVPAAP
AAALRPBPVDDGPTGPDVEDAADPTDVPDPELLRYLLGLRLTSGSEPEAAPRL
BRAVDODLGPVEVPENVLGALLRVKRLSNSPQAPARLLPP"
22. .120
sig_peptide
3'UTR      805. .945
polyA_signal
polyA_site 921. .926
BASE COUNT 134 a 355 c 307 g 149 t
ORIGIN
Query Match 59.68; Score 583; DB 10; Length 945;
Best Local Similarity 77.5%; Pred. No. 6e-60;
Matches 714; Conservative 1; Mismatches 205; Indels 1; Gaps 1;
QY 37 RGCCCGCTCGCTGGGAGCATGCGGGTGCCTGCTGCTGGGGCCGCGGCGGG 96
Db 1 AGCCCGCTCGTGGGAGCATGCGGGTGCCTGCTGCTGGGGCCGAGGCGGG 60
QY 97 GGCGTCGGGCTTTTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 156
Db 61 GGCGTAGGCTTTTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120
QY 157 GCGGCGCGGTAAAGAGACCCCGGGCTTAAGCGCAGGCTTCGCCCTTGAGCT 216
Db 121 GCGAGGCGCGTGAAGAACCCGCGAGTCTGAGCGCAGCATCCGCGCCCTT 180
QY 217 GGCGCTCTCCGCTTCCGCGCTGAGTGCCTGAGTGCCTGAGGCGGGGCGGT 276
Db 181 AGCACTCCCTCCCTGCTGCTGCGGCGTGCCTGAGGAGGCGGGTGGCGT 240
QY 277 GAGCTGCGGCGGCGCTGCGCATCTGCTGAGGCGCGAAGCTCAGGAGCGGCGG 336
Db 241 GAGCTGCGGCGGCGCTGCGCACCTGCTGAGGCGCGAGAGACAGAGCGCGT 300
QY 337 GAGCGCAGGAGGCTGAGGATCAGCAGCGCGGCTGCTGCGGCGAGCTGCTG 396
Db 301 GAGCGCAGGAGGCGGAGGATCAGCAGCGCGGCTGCTGCGCAGAGCTGCT 360
QY 397 GCGCGCCCGCCCAACTCTGATCCGGCTTGGGCTTGGACGACGACCCCGCG 456
Db 361 GGCTCCCGCGGCTGCTGCGACCCCGCTTGGCCCGGATGACGACCGGAGCT 420
QY 457 GCGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 516
Db 421 GCACAGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480
QY 517 CTTCCTCCCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 576
Db 481 CTTCCTCCCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
QY 577 GCGCGCGCGGCGGATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 636
Db 541 GCGCGCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 600
QY 637 CTGTTGAGGTACTTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 696
Db 601 CTGCTGAGGTACTTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 660
QY 697 GCGCGCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 756
```

```
Db 661 GCGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
QY 757 GCGGTGCTGGGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 816
Db 721 AACGTACTGGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
QY 817 GCAGCGCGGCTTTCGCCACCTGAGCACTGCGCGGATCCCGTGCACCTGG 876
Db 781 GCAGCGCGGCTTTCGCCACCTGAGCGCT-ACTACATTCTGCAGCCCTGG 839
QY 877 GTGCGCGCGGCTTTCGCCACCTGAGCACTGCTGCTGCTGCTGCTGCTG 936
Db 840 GCGCGCGGAGCAACCTGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 899
QY 937 CCGCGCGGAGCGGCTTCTCAC 957
Db 900 CCATCCCGCTGAGCGCTGCTGCT 920

RESULT 7
LOCUS BC012263 975 bp mRNA linear ROD 07-AUG-2002
DEFINITION Mus musculus, Similar to granin-like neuroendocrine peptide precursor, clone MGC:19107 IMAGE:4207854, mRNA, complete cds.
ACCESSION BC012263
VERSION BC012263.1 GI:15126669
KEYWORDS MGC.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Strausberg, R.
Direct Submission
Submitted (06-AUG-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgaps-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILNL at: http://image.llnl.gov
Series: IRAC Plate: 24 Row: p Column: 24.
Location/Qualifiers
FEATURES
source
1. .975
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="FVB/N"
/clone="MGC:19107 IMAGE:4207854"
/tissue_type="Colon, normal. 5 month old male mouse."
/clone_lib="NCI_CGAP_Co24"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"
20. .796
/codon_start=1
/product="Similar to granin-like neuroendocrine peptide precursor"
/protein_id="AAH12263.1"
/db_xref="GI:15126670"

CDS
```



```
/translation="MAGSPLLCGPRAGVGILVLLGLLRLPPTLSARPVKEPSLS
AASAPLVETSTPLRLRAVPGEAAGAVQELARALHLEAERQERARAEADQDQ
ARVLAQLRWGSPRASDPPLAPDDPDAPAAQALRALRLARLPAALAAQVLVPAAP
APRPFPYDGGPTGPDVEDAGDETPDVPDPELLRYLLGRIITGSSEPAAPARLRL
SYVDOLGPEVPPENIVGALLRVKLEPNFSPOAPARRLLPP"
BASE COUNT      173 a  353 c  304 g  145 t
ORIGIN

Query Match      59.3%; Score 580.2; DB 10; Length 975;
Best Local Similarity 80.0%; Pred. No. 1.3e-59;
Matches 726; Conservative 0; Mismatches 163; Indels 18; Gaps 3;

QY 39 CCGCGCTCGTGGGCGAGCATGGCGGCTCGCGCTGCTTGGGGCCCGCGGCGGGG 98
DB 1 CCGCGCTCGTGGGCGAGCATGGCGGCTCGCGCTGCTTGGGGCCCGCGGCGGGG 60
QY 99 CGTGCCTTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 158
DB 61 CGTGCCTTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120
QY 159 GCGGCGGTAAAGGACCCCGGCTTAAGCGAGGCTCTCGCCCTTGGCTGAGACTGG 218
DB 121 GAGGCGGTGAAGGACCCCGGCTTAAGCGAGGCTCTCGCCCTTGGCTGAGACTGG 180
QY 219 CGTCTCTCGCGCTTCCGCGGCTCACTGCGCGAGTGAAGCGGCGGCGGTGCGAG 278
DB 181 CACTCCCTCCGCTTCCGCGGCTGCGCGGAGGAGGCGGCGGTGCGGTGCGAGG 240
QY 279 GCTGCGCGGCGCTGCGCATCTGCTGAGGCGCAACGTCAAGAGCGGCGGCGCGCA 338
DB 241 GCTGCGCGGCGCTGCGCATCTGCTGAGGCGCAACGTCAAGAGCGGCGGCGCGCA 300
QY 339 GGGCGAGGAGGTGAGGATCAGAGCGCGGCTGCTGCGCGAGCTGCTGCGGCTTGGGG 398
DB 301 GGGCGAGGAGGTGAGGATCAGAGCGCGGCTGCTGCGCGAGCTGCTGCGGCTTGGGG 360
QY 399 CGCGCCCGCACTGATCCGCTGCTGCGCTTGGAGCGAGCACCGCGGCTGCTGCG 458
DB 361 CTCTCGCGTGTCTGCGAGCGCGCTTGGCGCGGAGATGACCGGAGCTCCAGCTG 420
QY 459 GCAGCTGCTGCGGCTGCTGCTGCGCGCGCTGACCTGCGCGCTAGCAGCGCGAGCT 518
DB 421 ACAGCTGCGCGGCTGCTGCTGCGAGCTGCGCTAGACCGCGCGCGCTGCGCGCAACT 480
QY 519 TGTCTCCGCGCGCTGCTGCGCGCGGCTGCGAGCGCGCGCGCGCGCTGACAGCGAG 578
DB 481 TGTCTCCGCG-----CCCTGCGGCTGCGCGCGAGCGCGCGCGCGCGCTGATGATG 534
QY 579 CCGCGCGCGCGGATGCTGAGGAGCGAGCGAGCGAGCACCGAGCTGAGCGCGAGCT 638
DB 535 CCGCGCTGCGCGAGAGCTGAGGATGCGCGGAGAGAGCTGCTGAGCTGAGCGCTGAGCT 594
QY 639 GTTAGGTACTTGTCTGGGACGATTTTGGGGAAGCGGAGATCCGAGGGGGTGGCAGC 698
DB 595 GCTGAGGTACTTGTAGGCGGATCTCACCAGGATTCGGAGCGAGAGCTGCTCTGCG 654
QY 699 CCGCGCGCGCTGCGCGCTGCGCGAGCGAGCTGAGCTGCGCGCTGAGGG 758
DB 655 CCGCGCGCGCTGCGCGGATCTGTGAGCAGGATTTGGGTGCGCGAGGTGCGCGCTGAGAA 714
QY 759 CPTGCTGGGGGCTGCTGCTGCTGTAAGCGCTAGAGACCGCGCGCGCGCGCTGCTG 818
DB 715 CGTACTGGGGGCTGCTGCTGCTGTAAGCGCTAGAGACCGCGCGCGCGCGCGCGCG 774
QY 819 AGCGCGCTCTTGGCAACCTGAGCACTGCGCGGATCCCGTGCACCTGGGACCGAGAGT 878
DB 775 AGCGCGCTCTTGGCAACCTGAGCGCTG-CTGATCTGACGCGCTTGAACCCAGGAGC 833
QY 879 GCGCGCGCATCCGCGCACAGGACTGCTGCGCGCGCGAGCGAGCTGCGAGCAACTTACCC 938
DB 834 GCGCGCACCAACCTG-----ACTCCCTGCGAGCAGCTGCAAGGCTGCTTACCC 882
QY 939 GCGCAGC 945
```

```
Db 883 AGCAACC 889
RESULT 8
AF293356
LOCUS      AF293356
DEFINITION Mus musculus IA4 mRNA, complete cds.
ACCESSION AF293356
VERSION    AF293356.1 GI:9954415
KEYWORDS
SOURCE     Mus musculus.
ORGANISM  Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1025)
Donadel,G., Marinos,N., Desilva,M.G., Lu,J., Notkins,A.L. and
Lan,M.S.
Molecular cloning and characterization of a highly basic protein,
IA-4, expressed in pancreatic islets and brain
Neuroendocrinology 67 (3), 190-196 (1998)
9630436
REFERENCE 2 (bases 1 to 1025)
AUTHORS   Donadel,G., Notkins,A.L. and Lan,M.S.
DIRECT SUBMISSION
Submitted (03-AUG-2000) Laboratory of Oral Medicine, National
Institute of Dental Research, 9000 Rockville Pike, Building 30 Room
124, Bethesda, MD 20892, USA
FEATURES
source
1..1025
Location/Qualifiers
/oranism="Mus musculus"
/db_xref="taxon:10090"
/cell_line="insulinoma b-TC-1"
94..657
/note="highly basic protein; highly expressed in
pancreatic islets and neuronal cells"
/codon_start=1
/product="IA4"
/db_xref="GI:9954416"
/translation="MAGSPLLCGPRAGVGILVLLGLLRLPPTLSARPVKEPSLS
ARSAPLVETSTPLRLRAVPGEAAGAVQELARALHLEAERQERARAEADQDQ
ARVLAQLRWGSPRASDPPLAPDDPDAPAAQALRALRLARLPAALAAQVLVPAAP
APRPFPYDGGPTGPDVEDAGDETPDVPDPELLRYLLGRIITGSSEPAAPARLRL
SYVDOLGPEVPPENIVGALLRVKLEPNFSPOAPARRLLPP"
BASE COUNT      156 a  382 c  330 g  157 t
ORIGIN

Query Match      51.3%; Score 502.2; DB 10; Length 1025;
Best Local Similarity 78.0%; Pred. No. 1.9e-50;
Matches 732; Conservative 5; Mismatches 173; Indels 29; Gaps 10;

QY 15 GCAGCAGCTCCGCCGCGMCGRRGCCGCTCGCTGGGCGAGCATGGCGGGTCCGCGT 74
DB 51 GACAGCAGCTCCGCCGCTCCGAGCGCGGCTCGTTGGGCGAGCATGGCGGGTCCGCGT 110
QY 75 GCTCTGGGCGCGCGCGCGGCGCTCGCGCTTTTGGTCTGCTGCTGCTGCGCGCTGT 134
DB 111 GCTCTGGGCGCGCGCGCGGCGCTCGGCATTTTGGTCTGCTGCTGCTGCGCGCTTCT 170
QY 135 TCGGCGCGCGCGCGCGCTCTGCGCGCGCGGCTAAAGAGAGCGCGCGCGCTAAGCGCAGC 194
DB 171 GAGGCTGCGCGCGCGCGCTGTCAGCGAGGCGCGGTGAAGAGAGCGCGCGAGTGTGAGCGCAGC 230
QY 195 GTCTCGCGCTTGGCTGAGACTGCGCGCTCTCGCGCGCTTCCGCGCGCTGAGTGCCTGAGG 254
DB 231 ATCCGCGCGCTTGGTTGAGCAGGAGCTCCCTCCCTTGGCTGCGCGCGCTGCGCGCGG 290
QY 255 TGAGCGCGCGCGGCGGCTG-CAGGAGCTGCGCG-----GGCGCTGCGCGCATCTGCTGAGAG 310
DB 291 AGAGCGCGCGGCTGCGGTGCGAGGAGCTGCGCGCGCGGCGCTGCGCGAGCTGCTGAGAG 350
QY 311 CCGAAGCTCAGGAGCGCGCGCGCGCGCGCGCGAGGCTGAGGATCAGAGCGCGCGCG 370
```

```

Db 351 CCGA--GACAGGAACCGCGCGTGTAGGCGCAGGAGGCTGAGGATCAGAGCGCATG 408
Qy TCCTGGCGCAGCTGCTGGCGCTCTGGGCGCCCGCCGCAACTCTGATCCG-GCTCTGGCG 429
Db TCCTGGCGCAGCTGCTGGCGCCCTGGGCGCTCTGGGCGCTCGGACCGCGCCCTTGGCC 468
Qy TTGGACGACGACCCGCGCCTGCGACGCGCTGCGAGCTGCGGCTCTGCTCTCGGCGCCGC 489
Db CCGGACGATGACCCGGGACGCTCCAGCTGTCACAGCTGCGCCGCTGCTCTCGGAGCTCGC 528
Qy CTTGACCTGCGCGCTAGCAGCCAGCTTGTCCCGCGCCCGCTCCCGCGCGCGCGCTC 549
Db CTAGACCGCGCGCGCTGGGGGCGCCAACTTGTCCCGC-----CCCTGCGCCCTGCGCGC 582
Qy CGACCCCGCGCGCGCTGTACGACGA-CGGCGCGCGCGCGCGGATGCTGAGGAGGCGAG 608
Db CGACCCCGCGCGCGCTGTATGATGATGATGGGCGCCACCTGGCCACAGCTCGAGGATGCGG 642
Qy CGACGAGACACCCGAGCTGGACCCGAGCTTTGAGGTACTTGTGGACGAGATCTTTC 668
Db CGACGAGACTGGTGACGT-GACCCCTGAGCTGCTGAGGTACTTGTAGGGCGGATCCTCAC 701
Qy GGGAGCGCGGACTCCGAGGGGTGGCAGCCCGCGCGCGCTCCCGCTGCGCGCGGACCA 728
Db GGAAGTTCGGAGGCGCAGAGGCTGTCTCTGCGCGCGCGCGCTCCCGGATCTGTGGACCA 761
Qy CGATGTGGGCTCTGAGTGTCCCGCTGAGGG--CGTGTGGGGCGCTGCTGCTGTGAA 786
Db GGAATTTGGGCGCGAGGTGCGCCCTGAGAGAACCTACTGGGGGTCTGCTACGCGTCAA 821
Qy GCGCTAGAGACCCCGCGCGCGCGCGCTGCTGCGCGCGCTCTGCGCGCGCTGCGCGCTG 846
Db GCGCTGAGAACCCCTGCGCGCGCGCGCGCGCGCGCTCTGCGCGCGCTGCGCGCGCTG 881
Qy CCGGATCCCGTGCACCTGGGACCCAGAGTGGCGCGCGCGCTCCCGCGCGCGCGAGATGC 906
Db -CTGATCTCTGACCGCGCTGGAACCCAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG-----AC 929
Qy TCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 945
Db TCCCTGCCAGCAGCTCCAGGCGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 968

RESULT 9
AF196971/c
LOCUS
DEFINITION
Homo sapiens GATA-binding protein 1 and histone deacetylase-like
protein genes, complete cds; CRAS pseudogene, complete sequence;
and protein translocase gene, partial cds.
ACCESSION
AF196971
VERSION
AF196971.1 GI:6289080
KEYWORDS
HTG.
SOURCE
Homo sapiens.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 113853)
Bleischmidt,K., Nyakatura,G., Strom,T.M., Drescher,B., Menzel,U.,
Weindl,A. and Rosenthal,A.
Direct Submission
Submitted (14-Oct-1999) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
COMMENT
1-36552: cosmid J1045; 11106-52106: cosmid E2237; 27017-67034:
cosmid L1850; 58536-92644: cosmid D1425; 80573-113853: cosmid
E1239.
FEATURES
source
1. .113853
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="Xp11.23"
/clone="cosmids J1045; E2237; L1850; D1425; E1239"
```

```

repeat_region
/evidence=not_experimental
/rpt_family="AluSx"
254..396
/evidence=not_experimental
/rpt_family="L1M4"
397..704
/evidence=not_experimental
/rpt_family="AluSg"
705..717
/evidence=not_experimental
/rpt_family="L1M4"
817..1083
/evidence=not_experimental
/rpt_family="L1MB7"
complement(1084..1371)
/evidence=not_experimental
/rpt_family="AluJo"
1372..1790
/evidence=not_experimental
/rpt_family="L1MB7"
1802..1924
/evidence=not_experimental
/rpt_family="L1MD2"
complement(2269..2440)
/evidence=not_experimental
/rpt_family="L2"
complement(2950..3059)
/evidence=not_experimental
/rpt_family="MIR"
3105..3235
/evidence=not_experimental
/rpt_family="MIR"
3268..3472
/evidence=not_experimental
/rpt_family="AluJo"
3504..3608
/notes="homology = 80.00%, score = 28, counts = 3"
/evidence=not_experimental
/rpt_type=tandem
/rpt_unit=ttagcattaggaatgatttactataataaattagta
3613..3955
/evidence=not_experimental
/rpt_family="L1MA5"
3964..4059
/notes="homology = 77.10%, score = 20, counts = 3"
/evidence=not_experimental
/rpt_type=tandem
/rpt_unit=aataataattagtagtattactataatcagtat
complement(4121..4369)
/evidence=not_experimental
/rpt_family="MIR"
4393..4463
/evidence=not_experimental
/rpt_family="L2"
complement(4502..4544)
/evidence=not_experimental
/rpt_family="MIR"
5062..5157
/notes="homology = 67.70%, score = 32, counts = 48"
/evidence=not_experimental
/rpt_type=tandem
/rpt_unit=tc
5649..5691
/evidence=not_experimental
/rpt_family="MIR"
6131..13747
/genes="GATA-1"
join(6131..6168,10614..10852,11367..11744,11846..11991,
12695..12820,13316..13747)
/genes="GATA-1"
/product="GATA-binding protein 1"
/notes="mRNA acc. no. M30601"
```

```
exon      6131..6168
/feature="GATA-1"
/number=1
repeat_region
/feature="not_experimental"
/feature="AluSx"
repeat_region
9694..9979
/feature="not_experimental"
/feature="AluJb"
10614..10852
/feature="GATA-1"
/number=2
CDS
join(10633..10852,11367..11744,11846..11991,12695..12820,
13316..13687)
/feature="GATA-1"
/feature="globin transcription factor 1; GF-1"
/codon_start=1
/product="GATA-binding protein 1"
/protein_id="AF06806.1"
/db_xref="GI:6289081"
/translation="MEFPLGLSLGHTSEPLPQFVDPALVSVSTPESGVFFPSGPEGLDAA
ASSTAPATAAALAYRYDAEAYRHSPVQVYPLINCMGIPGSGSPYAGWAYKVTG
LYPASTVCPTRDSPQAVEDLDGKSTSLFTLTKTERLSPDLLTLGPALPSLPVPTG
SAYGPDFSTFSPGSPPLNSAAYSSPKLGTLPPLPCAEKRCVNCATATPLMRD
RTHYLNCAGLIYHKMNGONRPLIRPKRLIYSKRAGTCTNCTTTTLWRNMSGD
PVCNAGLIYKHQVNRPLMRKDGITQIRNRKASGKRRKSGSLGSGTGAAGGAGG
MYVAGSGSGNCGEASGLTLGPPGTAHLIYGLGPVVLSGPVSHLMPFPGLGSPGTG
SPPTGMPPTTSTTVVAPLSS"
11367..11744
/feature="GATA-1"
/number=3
11846..11991
/feature="GATA-1"
/number=4
repeat_region
complement(12330..12477)
/feature="not_experimental"
/feature="MER5B"
12695..12820
/feature="GATA-1"
/number=5
13316..13747
/feature="GATA-1"
/number=6
repeat_region
14173..14226
/feature="homology = 98.10%, score = 25, counts = 2"
/feature="not_experimental"
/feature="tandem"
/feature="tandem"
14290..14375
/feature="homology = 62.80%, score = 20, counts = 43"
/feature="not_experimental"
/feature="tandem"
/feature="tandem"
complement(14876..14972)
/feature="not_experimental"
/feature="MER1B"
complement(14989..15178)
/feature="not_experimental"
/feature="MER1B"
16160..16259
/feature="homology = 74.00%, score = 23, counts = 4"
/feature="not_experimental"
/feature="tandem"
16338..16674
/feature="not_experimental"
/feature="MER1B"
16780..16856
/feature="homology = 86.00%, score = 22, counts = 3"
/feature="not_experimental"
/feature="tandem"
/feature="tandem"
16953..17048
```

```
/note="homology = 75.00%, score = 24, counts = 4"
/feature="not_experimental"
/feature="tandem"
/feature="tandem"
complement(17629..17940)
/feature="not_experimental"
/feature="MER33"
18028..18353
/feature="not_experimental"
/feature="MLT1"
complement(18388..18668)
/feature="not_experimental"
/feature="AluSx"
complement(18674..18977)
/feature="not_experimental"
/feature="AluSg"
19143..19448
/feature="not_experimental"
/feature="AluJb"
19750..19923
/feature="not_experimental"
/feature="MER58C"
complement(20505..20587)
/feature="not_experimental"
/feature="MIR"
20598..20658
/feature="not_experimental"

Query Match      48.4%; Score 474.2; DB 9; Length 113853;
Best Local Similarity 99.4%; Pred. No. 6.9e-48;
Matches 476; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 170 AGGAGCCCCCGCGCTAAGCGACGCGTCTCCGCCCTTGGCTGAGACTGGCGCTCTCGCC 229
    |||||
DB 51869 AGGAGCCCCCGCGCTAAGCGACGCGTCTCCGCCCTTGGCTGAGACTGGCGCTCTCGCC 51810

QY 230 GCTTCGCGCGGTGAGTCCCGCGAGGTGAGCGCGCGCGGGGGGGTGCAGGAGCTGGCGGGG 289
    |||||
DB 51809 GCTTCGCGCGGTGAGTCCCGCGAGGTGAGCGCGCGGGGGGGTGCAGGAGCTGGCGGGG 51750

QY 290 CGCTGGCGCATCTGCTGAGCGCGCAAGCTCAGGAGCGCGCGCGCGCGCGCGCGCGAGG 349
    |||||
DB 51749 CGCTGGCGCATCTGCTGAGCGCGCAAGCTCAGGAGCGCGCGCGCGCGCGCGCGAGG 51690

QY 350 CTGAGGATCAGCAGCGCGCTCTGCGCGAGCTGCTGCGCGTCTGGGGGGCGCGCGCGCGCA 409
    |||||
DB 51689 CTGAGGATCAGCAGCGCGCTCTGCGCGAGCTGCTGCGCGTCTGGGGGGCGCGCGCGCA 51630

QY 410 ACTCTGATCGCGCTGCGGTGAGCAGCAGCAGCGCGCGCGCGCGCGCGCGCGCTCGCTC 469
    |||||
DB 51629 ACTCTGATCGCGCTGCGGTGAGCAGCAGCAGCGCGCGCGCGCGCGCGCGCTCGCTC 51570

QY 470 GCGCTCTGCTCCGCGCGCGCTTACCTCGCGCGCTAGCAGCGCGCGCGCGCGCGCGCG 529
    |||||
DB 51569 GCGCTCTGCTCCGCGCGCGCTTACCTCGCGCGCTAGCAGCGCGCGCGCGCGCGCGCG 51510

QY 530 CCGTCCCGCGCGCGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 589
    |||||
DB 51509 CCGTCCCGCGCGCGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 51450

QY 590 CGGATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 648
    |||||
DB 51449 CGGATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 51391

RESULT 10
AC098339
LOCUS
DEFINITION
Rattus norvegicus clone CH230-2A16, *** SEQUENCING IN PROGRESS ***
AC098339
AC098339.3 GI:22474728
VERSION
KEYWORDS
HTG; HTGS_PHASE1.
```

SOURCE
ORGANISM

Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS

1 (bases 1 to 62396)
Munzly D.M., Adams C., Adio-Oduola B., Ali-osman, F.R., Allen C.,
Albrechts S.L., Amarantunge H.C., Are J.R., Ayele M., Banks T.,
Barbala J., Benton J., Blmage K., Blankenburg K., Bonnin D.,
Bouck J., Bowie S., Brileva M., Brown E., Brown M., Bryant N.P.,
Buhay C., Burch P., Burkett C., Burrell K.L., Byrd N.C.,
Caron T.F., Carter M., Cavazos S.R., Chacko J., Chavez D.,
Chen G., Chen R., Chen Z., Chowdhry I., Christopoulos C.,
Cleveland C.D., Cox C., Coyle M.D., Dathorne S.R., David R.,
Davila M.L., Davis C., Davy-Carroll L., Dederich D.A.,
Delaney K.R., Delgado O., Denn A.L., Ding Y., Dinah H.H.,
Douthwaite K.J., Draper H., Dugan-Rocha S., Durbin K.J.,
Earhart C., Edgar D., Edwards C.C., Elhaj C., Escotto M.,
Falls T., Ferraguto D., Flagg N., Ford J., Foster P., Frantz P.,
Galsi A., Gao J., Garcia A., Garner T., Garza N., Gill R.,
Gorrell J.H., Guevara W., Gunaratne P., Hale S., Hamilton K.,
Harris C., Harris K., Hart M., Havlak P., Hawes A., Hernandez J.,
Hernandez O., Hodgson A., Hogue M., Holloway C., Hollins B.,
Honsi F., Howard S., Huber J., Hulyk S., Hume J., Jackson L.E.,
Jacobson B., Jia F., Johnson R., Jolivet S., Joudah S.,
Karlisson E., Kelly S., Khan U., King L., Korvah J., Kovar C.,
Kratovic J., Kureshi A., Landry N., Leal B., Lewis L.C., Lewis L.,
Li J., Li Z., Lichtarge O., Lieu C., Liu J., Liu W., Loulseged H.,
Lozado R.J., Lu X., Lucier A., Lucier R., Luna R., Ma J.,
Maheshwari M., Mapua P., Martin R., Martindale A., Martinez E.,
Massey E., Mahoney E., McLeod M.P., Meador M., Mei G., Metzker M.,
Miner G., Miner Z., Mitchell T., Mohabbat K., Morgan M., Morris S.,
Mosier M., Neal D., Newton J., Newton N., Nguyen A., Nguyen N.,
Nguyen N., Nickerson E., Nwokenwo S., Ogih M., Okwuonu G.,
Oraguine N., Oviedo R., Pace A., Payton B., Peery J., Perez L.,
Peterson L., Pickens R., Primus E., Pu L.L., Quiles M., Ren Y.,
Rivers M., Rojas A., Rojokan I., Rolfe M., Ruiz S., Savery G.,
Scherer S., Scott G., Shen H., Shoohtari N., Sison I.,
Sodergren E., Sonaite T., Sparks A., Stanley H., Stone H.,
Sutton A., Svatek A., Tabor P., Tamerisa A., Tamerisa K., Tang H.,
Tantey J., Taylor C., Taylor T., Telford B., Thomas N., Thomas S.,
Usmani K., Vasquez L., Vera V., Villalon D., Vinson R., Wang Q.,
Wang S., Ward-Moore S., Warren R., Washington C., Watlington S.,
Williams G., Williamson A., Wleczky R., Wooden S., Worley K.,
Wu C., Wu Y., Wu Y.F., Zhou J., Zorrilla S., Nelson D.,
Weinstock G. and Gibbs R.

TITLE
JOURNAL
REFERENCE
AUTHORS

Direct Submission
2 (bases 1 to 62396)
Worley K.C.
Submitted (23-OCT-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS
TITLE
JOURNAL

3 (bases 1 to 62396)
Worley K.C.
Direct Submission
Submitted (24-AUG-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Aug 24, 2002 this sequence version replaced gi:21953722.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: TULN
Center clone name: CH230-2A16
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329

Consensus quality: 23320 bases at least Q40
Consensus quality: 26055 bases at least Q30
Consensus quality: 27202 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 38 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1088: contig of 1088 bp in length
* 1089: contig of unknown length
* 1189: contig of 1144 bp in length
* 2332: contig of 1205 bp in length
* 2432: gap of unknown length
* 3637: contig of 1205 bp in length
* 3638: gap of unknown length
* 3738: contig of 1147 bp in length
* 4885: gap of unknown length
* 4885: gap of unknown length
* 6231: contig of 1247 bp in length
* 6232: gap of unknown length
* 6331: contig of 1066 bp in length
* 7398: gap of unknown length
* 7498: contig of 1105 bp in length
* 8603: gap of unknown length
* 8702: contig of 1006 bp in length
* 8703: contig of unknown length
* 9709: gap of unknown length
* 9809: contig of 1287 bp in length
* 11096: gap of unknown length
* 11196: contig of 1394 bp in length
* 12590: gap of unknown length
* 12690: contig of 1178 bp in length
* 13868: gap of unknown length
* 13968: contig of 1183 bp in length
* 15151: gap of unknown length
* 15251: contig of 1141 bp in length
* 16392: gap of unknown length
* 16492: contig of 1037 bp in length
* 17529: gap of unknown length
* 17629: contig of 1278 bp in length
* 18907: gap of unknown length
* 19007: contig of 1146 bp in length
* 20153: gap of unknown length
* 20252: contig of 1040 bp in length
* 21292: gap of unknown length
* 21393: contig of 1308 bp in length
* 22701: gap of unknown length
* 22801: contig of 1906 bp in length
* 24707: gap of unknown length
* 24806: contig of 1582 bp in length
* 26389: gap of unknown length
* 26489: contig of 1149 bp in length
* 27638: gap of unknown length
* 27738: contig of 1365 bp in length
* 29103: gap of unknown length
* 29203: contig of 1150 bp in length
* 30353: gap of unknown length
* 30453: contig of 1268 bp in length
* 31721: gap of unknown length
* 31820: contig of 2083 bp in length
* 33903: gap of unknown length
* 34003: contig of 1756 bp in length
* 35759: gap of unknown length
* 35859: contig of 1381 bp in length
* 37240: gap of unknown length
* 37341: contig of 1882 bp in length
* 39223: gap of unknown length
* 39323: contig of 1601 bp in length
* 40924: gap of unknown length
* 41024: contig of 1859 bp in length

```

* 42883 42982: gap of unknown length
* 42983 44647: contig of 1665 bp in length
* 44648 44747: gap of unknown length
* 44748 46546: contig of 1799 bp in length
* 46547 46547: gap of unknown length
* 46547 48461: contig of 1815 bp in length
* 48462 48561: gap of unknown length
* 48562 51223: contig of 2662 bp in length
* 51224 51324: gap of unknown length
* 51324 53891: contig of 2568 bp in length
* 53892 53991: gap of unknown length
* 53992 56351: contig of 2360 bp in length
* 56352 56451: gap of unknown length
* 56452 58948: contig of 2497 bp in length
* 58949 62396: gap of unknown length
* 59049 62396: contig of 3348 bp in length.
FEATURES
    source
    1. .62396
    /organism="Rattus norvegicus"
    /db_xref="taxon:10116"
    /clone="CH230-2A16"
BASE COUNT 15738 a 13739 c 13230 g 15967 t 3722 others
ORIGIN

Query Match      35.0%; Score 343; DB 2; Length 62396;
Best Local Similarity 82.3%;
Matches 394; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 170 AGGAGCGCCGCGCGCTAAGCGCAGCGCTCCGCCCTTGGCTGAGACTGGCGCTCCCTCGCC 229
Db 54162 AGGAACCCCGCAGCTGAGCGCGAGCATCCGCCCTTGGCTGAGAGCAGCTCCCTCC 54221

QY 230 GCTTCCGCGCGCTCAGTGTCCCGAGGTGAGCGCGCGCGCGGTGTCAGGAGCTGGCGCGG 289
Db 54222 GCTTGGCTCGCGCGCTGCCCGGAGGAGCGCGCGGTGCGGTGTCAGGAGCTGGCGCGG 54281

QY 290 CGCTGGCGCATCTGCTGGAGCGCGAAGCTCAGGAGCGCGCGCGCGCGCGCGCGCGAGG 349
Db 54282 CGCTGGCGCATCTGCTGGAGCGCGCGAGAGCAGGAGCGCGCGCGCGCGCGCGCGCGAGG 54341

QY 350 CTGAGGATCAGCAGCGCGCGCTCTCGCGCAGCTGCTGCGCGCTGCGCGCGCGCGCGCGCA 409
Db 54342 CGGAGGATCAGCAGCGCGCGCTCTGCGCAGCTGCTGCGCGCTGCGCGCTCCCGCGTG 54401

QY 410 ACTGTGATCGGCTGTGGGCTTGGAGCAGCAGCCCGCGCGCGCTGCGAGCGCAGCTCGCTC 469
Db 54402 CCTCGGACCGCGCTTGGCGCGGATGACGACCGCGCGCGCTCCCGCTGCGACAGCTCGCC 54461

QY 470 GCGCTGTGCTCCGCGCGCGCTTGACCTGCGCGCGCTAGCAGCGCGCGCTGTCCCGCGCG 529
Db 54462 GTGCTGTGCTCCGAGCTGCGCTGAGACCGCGCGCGCGCGCGCGCTGTCCCGCGCG 54521

QY 530 CCGTCCCGCGCGCGCGCTCCGACCGCGCGCGCGCTCTACGACGAGCGCGCGCGCGCGCC 589
Db 54522 CTGCGCTGCGCGCTGCGCTGCGAGCGCGCGCGCGCGCGCGCTATGATGATGCGCGCGCG 54581

QY 590 CGGATGCTGAGGAGCGCGCGCAGCAGCAGCCCGCGCGCGCGCGCGCGCTGTTCAGGTAC 648
Db 54582 CAGACGTCGAGGATGCGCGCTGAGCAGACTCCTGATGTGGACCGCTGCTGAGGTGC 54640

RESULT 11
AC098473
LOCUS
DEFINITION
AC098473
Rattus norvegicus clone CH230-2J23, *** SEQUENCING IN PROGRESS ***,
36 unordered pieces.
ACCESSION
AC098473
VERSION
AC098473.3 GI:21953422
KEYWORDS
HTG; HTGS_PHASE1.
SOURCE
Norway rat.
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae;
```

REFERENCE
AUTHORS

Rattus.
1 (bases 1 to 77876)
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
Alsbrooks, S.L., Amaral, H.C., Are, J.R., Ayale, M., Banks, T.,
Barbaria, J., Benton, J., Binage, K., Blankenburg, K., Bonnin, D.,
Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Coyle, M.D., Dathorne, S.R., David, R.,
Cleveland, C.D., Cox, C., Davy-Carroll, L., Dederich, D.A.,
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
Fallis, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
Gabi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Havlak, P., Hayes, A., Hernandez, J.,
Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B.,
Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loughheed, H.,
Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okunolu, G.,
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
Rivers, M., Rojas, A., Rojubokan, I., Roife, M., Ruiz, S., Savary, G.,
Scherer, S., Scott, G., Shen, H., Shoostari, N., Sisson, I.,
Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, R., Thomas, S.,
Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Wlezyk, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

2 (bases 1 to 77876)
Worley, K.C.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Submitted (23-Oct-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 77876)
Worley, K.C.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Submitted (31-Jul-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 24, 2002 this sequence version replaced gi:20976551.

COMMENT

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: TUVZ
Center clone name: CH230-2J23
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 42926 bases at least Q40
Consensus quality: 47276 bases at least Q30
Consensus quality: 50269 bases at least Q20

Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, K.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jollivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, E., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louised, H., Lozadó, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, F., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokoku, S., Ogih, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshitari, N., Sisson, I., Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmari, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G., and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 111948)
Worley, K.C.

Direct Submission
Submitted (15-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 111948)
Worley, K.C.

Direct Submission
Submitted (11-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jul 10, 2002 this sequence version replaced gi:17941324.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GAXL
Center clone name: CH230-4x5
----- Summary Statistics
Sequencing vector: plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 54807 bases at least Q40
Consensus quality: 60236 bases at least Q30
Consensus quality: 64720 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 53 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence

* as soon as it is available and the accession number will
* be preserved.

1 1001: contig of 1001 bp in length
1002 1101: gap of unknown length
1102 2332: contig of 1231 bp in length
2333 2432: gap of unknown length
2433 3485: contig of 1053 bp in length
3486 3585: gap of unknown length
3586 4814: contig of 1229 bp in length
4815 4914: gap of unknown length
4915 6154: contig of 1240 bp in length
6155 6254: gap of unknown length
6255 7917: contig of 1663 bp in length
7918 8017: gap of unknown length
8019 9057: contig of 1040 bp in length
9058 9157: gap of unknown length
9158 10364: contig of 1207 bp in length
10365 10464: gap of unknown length
10465 11897: contig of 1433 bp in length
11898 11997: gap of unknown length
11999 13263: contig of 1266 bp in length
13264 13363: gap of unknown length
13364 14544: contig of 1181 bp in length
14545 14644: gap of unknown length
14645 16064: contig of 1420 bp in length
16065 16164: gap of unknown length
16165 17785: contig of 1621 bp in length
17786 17885: gap of unknown length
17886 19123: contig of 1238 bp in length
19124 19223: gap of unknown length
19224 20631: contig of 1408 bp in length
20632 20731: gap of unknown length
20732 22214: contig of 1483 bp in length
22215 22314: gap of unknown length
22315 23749: contig of 1435 bp in length
23750 23849: gap of unknown length
23850 25423: contig of 1574 bp in length
25424 25523: gap of unknown length
25524 26850: contig of 1327 bp in length
26851 26950: gap of unknown length
26951 28487: contig of 1537 bp in length
28488 28588: gap of unknown length
28589 29836: contig of 1249 bp in length
29837 31672: contig of 1736 bp in length
31673 31772: gap of unknown length
31773 33161: contig of 1389 bp in length
33162 33262: gap of unknown length
33263 34438: contig of 1176 bp in length
34439 36752: contig of 2215 bp in length
36753 36852: gap of unknown length
36853 37909: contig of 1057 bp in length
37910 38009: gap of unknown length
38010 39170: contig of 1161 bp in length
39171 39270: gap of unknown length
39271 41040: contig of 1770 bp in length
41041 41140: gap of unknown length
41141 42441: contig of 1301 bp in length
42442 42541: gap of unknown length
42542 43610: contig of 1069 bp in length
43611 43710: gap of unknown length
43711 45019: contig of 1309 bp in length
45020 45119: gap of unknown length
45120 46358: contig of 1239 bp in length
46359 46458: gap of unknown length
46459 48595: contig of 2137 bp in length
48596 48696: gap of unknown length
48697 50326: contig of 1631 bp in length
50327 50426: gap of unknown length
50427 51957: contig of 1531 bp in length
51958 52057: gap of unknown length
52058 53961: contig of 1904 bp in length

```
* 53962 54061: gap of unknown length
* 54062 57027: contig of 2966 bp in length
* 57028 57127: gap of unknown length
* 57128 58760: contig of 1633 bp in length
* 58761 58860: gap of unknown length
* 58861 61388: contig of 2528 bp in length
* 61389 61488: gap of unknown length
* 61489 63296: contig of 1808 bp in length
* 63297 63396: gap of unknown length
* 63397 65907: contig of 2511 bp in length
* 65908 66007: gap of unknown length
* 66008 69562: contig of 3555 bp in length
* 69563 69662: gap of unknown length
* 69663 72572: contig of 2910 bp in length
* 72573 72672: gap of unknown length
* 72673 75269: contig of 2597 bp in length
* 75270 75369: gap of unknown length
* 75370 78347: contig of 2978 bp in length
* 78348 78447: gap of unknown length
* 78448 82487: contig of 4040 bp in length
* 82488 82587: gap of unknown length
* 82588 84648: contig of 2061 bp in length
* 84649 84748: gap of unknown length
* 84749 87601: contig of 2853 bp in length
* 87602 87701: gap of unknown length
* 87702 91380: contig of 3679 bp in length
* 91381 91480: gap of unknown length
* 91481 94959: contig of 3479 bp in length
* 94960 95059: gap of unknown length
* 95060 99432: contig of 4373 bp in length
* 99433 99532: gap of unknown length
* 99533 105199: contig of 5666 bp in length
* 105199 105299: gap of unknown length
* 105299 111948: contig of 6650 bp in length.
FEATURES
    source
    Location/Qualifiers
        1..111948
            Query Match      35.08; Score 343; DB 2; Length 111948;
            Best Local Similarity 82.3%; Pred. No. 1.9e-32;
            Matches 394; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

Qy 170 AGGAGCCCGCGGCTAGACGAGCTCGCCCTTGGCTGAGACTCGCGCTCTCGCC 229
Db 107300 AGGAGCCCGCGGCTAGACGAGCTCGCCCTTGGCTGAGACTCGCGCTCTCGCC 107241

Qy 230 GCTTCGCGGCTAGTCCCGGAGTGAGCGCGGCGGCGGTGTCAGAGCTGCGCGG 289
Db 107240 GCTTCGCGGCTAGTCCCGGAGTGAGCGCGGCGGCGGTGTCAGAGCTGCGCGG 107181

Qy 290 CGCTGGCGCCTGCTGGAGCGGAGACAGAGCGCGGCTGCTGAGCGCAGGAGG 349
Db 107180 CGCTGGCGCCTGCTGGAGCGGAGACAGAGCGCGGCTGCTGAGCGCAGGAGG 107121

Qy 350 CTGAGGATCAGCAGCGCGCTCTGCGGAGCTGCTGCGGCTGCGGCGCGCCCGCA 409
Db 107120 CTGAGGATCAGCAGCGCGCTCTGCGGAGCTGCTGCGGCTGCGGCGCGCCG 107061

Qy 410 ACTGTGATCCGCGCTTGGGCTTGGACGACGACCGCGGCTGTCAGCGCAGCTCGTC 469
Db 107060 CCTCGGACCGCGCTTGGGCTTGGACGACGACCGCGGCTGTCAGCGCAGCTCGTC 107001

Qy 470 CGCTCTGCTCCGCGCGCGCTTGAACCTTGCAGCGCTAGCAGCCAGCTTTCGCCGCG 529
Db 107000 GTGCTCTGCTCCGAGCTGCGCTAGACCGCGCGCGCTTGCAGCTTTCGCCGCG 106941

Qy 530 CCGTCCCGCGCGCGCTCCGACCGCGCGCGCTGTCAGCAGCGCGCGCGCGCGCG 589
Db 106940 CTGCGCGCTGCGCGCTGCGCGCGCGCGCGCGCGCTGTCAGCAGCGCGCGCG 106881

Qy 590 CGGATGCTGAGGAGCGGCGGACGAGACACCGGACGTGGACCGCGCGCTGTGAGGTAC 648
Db 106880 CAGACGTGAGGAGCGGCGGCTGACGAGACTCTCTGATGTGGACCGCTGAGCTGCTGAGGTGC 106822
```

AL670169 151694 bp DNA linear ROD 23-MAY-2002
Mouse DNA sequence from clone RP23-198C2 on chromosome X, complete sequence.

AL670169 GI:21213366
HTG.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 151694)

Direct Submission
Submitted (23-MAY-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
On May 25, 2002 this sequence version replaced gi:21104114.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-198C2 is from the RPCI-23 Mouse PAC Library constructed by the group of Pieter de Jong.
For further details see <http://www.chori.org/bacpac/home.htm>
VECTOR: PBACE3.6.

Location/Qualifiers
1..151694
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="X"
/clone="RP23-198C2"
/clone_lib="RPCI-23"

BASE COUNT 39029 a 35150 c 35380 g 42135 t

ORIGIN

Query Match 34.08; Score 332.6; DB 10; Length 151694;
Best Local Similarity 79.5%; Pred. No. 2.8e-31;
Matches 408; Conservative 0; Mismatches 99; Indels 6; Gaps 1;

Qy 168 AAAGAGAGCCCGCGGCTTAAGCGAGCTCTCCGCCCTTGGCTGAGAGTGGCGTCTCCG 227
Db 24119 ACAGAGAGCCCGCGAGTCTGAGCGAGCATCGCGCCCTTGTGAGAGGAGCACTCCCT 24178

Qy 228 CCGCTTCGCGGCTCAGTGCAGGCGGCGGCGGCGGCGGCTGTCAGGAGTGGCGCG 287
Db 24179 CCGCTTCGCTGCGGCGGCTGCGCGGAGGAGCGGCGGCTGCGGAGGAGTGGCGCG 24238

Qy 288 GGCCTGCGCATCTGCTGAGGCGGAGGAGTCAAGAGCGGCGGCGGCGGAGGCGCAGGA 347
Db 24239 GGCCTGCGCAGCTGCTGAGGCGGAGAGAGAGAGAGCGGCGGCTGCTGAGGCGCAGGA 24298

Qy 348 GGCTGAGGATCAGAGCGCGCGCTCTGGCGAGCTGCTGCGCGCTTGGGCGCGCGCGCG 407
Db 24299 GGCTGAGGATCAGAGCGCGCGCTGCTGGCGAGCTGCTGCGCGCTTGGGCGCTCTCCGCG 24358

Qy 408 CAACCTCTGATCCGCGCTCTGGCTTGGAGGAGAGACCCCGACCGCGCTGCGAGCGCAGCTCGC 467


```

Db 24359 TGCTGGAGCCCGCTTGGCCCGGACGATGACCGGAGCCTCCAGCTGCACAGCTCGC 24418
QY 468 TCAGCTCTCTCCGCGCCGCTTGGACCTGCGCCCTAGCAGCCAGCTTGTCCCGCC 527
Db 24419 CCCTGCTCTCTCCGAGCTCGCTAGACCGCGCGCCCTGGCAGCCCACTTGTCCCGC 24478
QY 528 GCCGTCCTCCCGCGCGCTTCCGACCCCGCGCGCCCTAGCAGACGCGCCCGCGGG 587
Db 24479 -----CCCTGCGCGCTGCGCGGACCCCGCGCGCCCTAGTATGATGCGCCCACTGG 24532
QY 588 CCGGATGCTGAGGAGCGGACGAGCAGACACCGGAGCTGACCGGAGCTGTGAGTA 647
Db 24533 CCAGAGCTCGAGATGCGCGGACGAGACTCTGAGCTGAGCGCCTGAGCTGCTGAGTG 24592
QY 648 TTCTGCTGGGACGATTTTTCGCGGAAGCGCGGA 680
Db 24593 CGGGACCAAGGGGAGGGAGGCAAGGGCAGA 24625

RESULT 14
AC123014
LOCUS
DEFINITION
Rattus norvegicus clone CH230-150E3, *** SEQUENCING IN PROGRESS
*** 63 unordered pieces.
AC123014
AC123014.2 GI:21909084
VERSION
HTG: HTGS_PHASE1.
KEYWORDS
Rattus norvegicus.
SOURCE
Rattus norvegicus
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 177613)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbarta,J., Benton,J., Bimager,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowle,S., Brivea,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhardt,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,
Homsí,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlssoen,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseghe,H.,
Lohado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mel,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okuwonu,G.,
Oraguyn,N., Oviedo,E., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I.,
Sodergren,E., Sonaite,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 177613)
Worley,K.C.
Direct Submission
Submitted (26-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 177613)
Worley,K.C.
Direct Submission
Submitted (24-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 19, 2002 this sequence version replaced gi:21217349.

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GYVN

Center clone name: CH230-150E3

----- Summary Statistics

Sequencing vector: Plasmid;

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 131005 bases at least Q40

Consensus quality: 135917 bases at least Q30

Consensus quality: 139886 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 63 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 1002: contig of 1002 bp in length

* 1003 1102: gap of unknown length

* 1103 2123: contig of 1021 bp in length

* 2124 2223: gap of unknown length

* 2224 3343: contig of 1120 bp in length

* 3344 3443: gap of unknown length

* 3444 4556: contig of 1113 bp in length

* 4557 4656: gap of unknown length

* 4657 5684: contig of 1028 bp in length

* 5685 5784: gap of unknown length

* 5785 7333: contig of 1549 bp in length

* 7334 7433: gap of unknown length

* 7434 8575: contig of 1142 bp in length

* 8576 8675: gap of unknown length

* 8676 10032: contig of 1357 bp in length

* 10033 10132: gap of unknown length

* 10133 11192: contig of 1060 bp in length

* 11193 11292: gap of unknown length

* 11293 12599: contig of 1307 bp in length

* 12600 12699: gap of unknown length

* 12700 13872: contig of 1173 bp in length

* 13873 13972: gap of unknown length

* 13973 15404: contig of 1432 bp in length

* 15405 15504: gap of unknown length

* 15505 16651: contig of 1147 bp in length

* 16652 16751: gap of unknown length

* 16752 17847: contig of 1096 bp in length

* 17848 17947: gap of unknown length

* 17948 19740: contig of 1793 bp in length

* 19741 19840: gap of unknown length

* 19841 21890: contig of 2050 bp in length


```

QY 763 CTGGGGCGCTGCTGCGTGTGAACGCTAGAGACCCCGCGCCCGAGGTGCTGCACGC 822
Db 241 CTGGGGCGCGTGTGCGTGTGAACGCTAGAGACCCCGCGCCCGAGGTGCTGCACGC 182
QY 823 CGCCTCTTGCCACCCCTGAGCACTGCCCCGGATCCCGTGCACCCCTGGGACCCAGAAAGTGCCCC 882
Db 181 CGCCTCTTGCCACCCCTGAGCACTGCCCCGGATCCCGTGCACCCCTGGGACCCAGAAAGTGCCCC 122
QY 883 CGGCCATCCCGCCACCCAGGACTGCTCCCGCCAGCAGCTCCAGAGCAACTTACCCCGGCC 942
Db 121 CGGCCATCCCGCCACCCAGGACTGCTCCCGCCAGCAGCTCCAGAGCAACTTACCCCGGCC 62
QY 943 AGCCAGCCCTCTCACCCGAGGATCCCTACCCCGCTGGC 979
Db 61 AGCCAGCCCTCTCACCCGAGGATCCCTACCCCGCTGGC 25

```

Search completed: April 12, 2003, 18:13:55
Job time : 3115 secs

